Neo-morphic Missense Mutant P53 Proteins and The Co-drivers Promoting Cell Invasion

by

Anasuya Pal

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2019 by the
Graduate Supervisory Committee:

Joshua LaBaer, Chair
Robert Roberson
Wade Van Horn
Carlo Maley

ARIZONA STATE UNIVERSITY

August 2019

ABSTRACT

Phenotypic and molecular profiling demonstrates a high degree of heterogeneity in the breast tumors. *TP53* tumor suppressor is mutated in 30% of all breast tumors and the mutation frequency in basal-like subtype is as high as 80% and co-exists with several other somatic mutations in different genes. It was hypothesized that tumor heterogeneity is a result of a combination of neo-morphic functions of specific *TP53* driver mutations and distinct co-mutations or the co-drivers for each type of *TP53* mutation. The 10 most common p53 missense mutant proteins found in breast cancer patients were ectopically expressed in normal-like mammary epithelial cells and phenotypes associated with various hallmarks of cancer examined. Supporting the hypothesis, a wide spectrum of phenotypic changes in cell survival, resistance to apoptosis and anoikis, cell migration, invasion and polarity was observed in the mutants compared to wildtype p53 expressing cells. The missense mutants R248W, R273C and Y220C were most aggressive. Integrated analysis of ChIP and RNA seq showed distinct promoter binding profiles of the p53 mutant proteins different than wildtype p53, implying altered transcriptional activity of mutant p53 proteins and the phenotypic heterogeneity of tumors. Enrichment and model-based pathway analyses revealed dysregulated adherens junction and focal adhesion pathways associated with the aggressive p53 mutants. As several somatic mutations co-appear with mutant *TP53*, we performed a functional assay to fish out the relevant collaborating driver mutations, the co-drivers. When *PTEN* was deleted by CRISPR-Cas9 in non-invasive p53-Y234C mutant cell, an increase in cell invasion was observed justifying the concept of co-drivers. A genome wide CRISPR library-based screen on p53-Y234C and R273C cells identified separate candidate co-driver mutations that promoted cell invasion. The top candidates included several mutated genes in breast cancer patients harboring *TP53* mutations and were associated with cytoskeletal and apoptosis resistance pathways. Overall, the combined approach of molecular profiling and functional genomics screen highlighted distinct sets of co-driver mutations that can lead to heterogeneous phenotypes and promote aggressiveness in cells with different *TP53* mutation background, which can guide development of novel targeted therapies.

ACKNOWLEDGMENTS

This extensive work has been possible with the support of many people at the Virginia G. Piper Center for Personalized Diagnostics (CPD), The Biodesign Institute at Arizona State University. I wish to express my deep gratitude to Dr. Joshua LaBaer for accepting me as a graduate student in the lab and giving me the wonderful opportunity to work on this project. Dr. LaBaer has been a constant support and I thank him for all the suggestions and constructive criticism of the research all throughout the years. I am also thankful to all the members of my supervisory committee, Dr. Roberson, Dr. Van Horn and Dr. Maley for reviewing my work and helping me in improving it from different aspects.

There are a few people who need a special mention as without them carrying out this work would have been either very difficult or impossible:

1) Dr. Laura Gonzalez – She vividly explained the whole project and initially trained me in various aspects of it. She also convinced Dr. LaBaer that I would be a good graduate student and believed in my abilities. Laura gave me a whole new perspective on how to approach any scientific problem.

2) Dr. Jin Park, Assistant Research Professor, The Biodesign Institute, ASU – He has immensely helped me in shaping up the whole project. I cannot thank him enough for helping me in the bioinformatic analysis and to make this study a comprehensive meaningful story. His positive energy makes things easy and achievable.

3) Seron Eaton, Grant Christensen and Chenxi Xu– Seron has been a great co-worker and a constant support all throughout. He trained me in mammalian cell culture and the cell-based assays that I have been doing all along my study. He has been a great friend and a problem solver. All the mice study would not have been possible without Grant. He has been the best student I ever had and helped me with the *in-vivo* studies even after his graduation from ASU. Chenxi is the person behind all the bioinformatic analysis and developed the wonderful figures for the thesis and papers.

Gene editing using the CRISPR-Cas9 system has been quite challenging and Jessica Lang helped me immensely to get it to work. Jacquelyn Kilbourne and Ken Lowe were helpful for the mice surgeries for developing the breast tumor and metastasis model. The DNASU sequencing team, Jason Steel, Nicholas Mellor and Joy Blain, have been wonderful to work with. I would like to thank Femina Rauf for being a very good friend and a great lab mate. She has helped me with the experiments from time to time and been a patient listener to my problems. I would also like to thank all the members of the lab for their help and support. Funding from the Breast Cancer Research Foundation (BCRF) has supported this research throughout the years.

My parents have been a huge source of encouragement and mental support. I cannot thank them enough for their belief in me. I honestly thank my husband Robin for supporting me in various ways during the difficult PhD life. It was tough to handle everything being physically apart. I am also grateful to my parents' in-law for encouraging me to pursue my dreams. Finally, I thank two of my best friends Arunima Coomar and Anshu Sharma for getting me through tough times and making life enjoyable.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

IDENTIFYING GENE CANDIDATES FOR DEVELOPING ALTERNATE TARGETED THERAPIES FOR AGGRESSIVE BREAST CANCER HARBORING DIFFERENT TP53 MISSENSE MUTATIONS: BACKGROUND, HYPOTHESIS AND APPROACHES

Introduction

Overview of human breast cancer

The breast cancer tumor evolves from a transformed normal cell via hyperplasia, premalignant tumor, in-situ carcinoma to invasive carcinoma by accumulation of various genetic alterations (Beckmann 1997). Consequently, breast cancer displays heterogeneity with inter and intra tumor diversity that influences breast cancer progression and therapy resistance (Polyak 2011). Gene expression analysis of breast tumors have identified the major subtypes- luminal, HER2 overexpressing and basal-like. Each subtype has different genetic constitution, risk factor, response to treatment, disease progression and pattern of metastasis. Luminal tumors are hormone receptor positive (ER+ and PR+) and are treated with hormonal therapies. HER2 amplified tumors respond well to anti-HER2 antibody trastuzumab. But there is no targeted therapy for basal-like subtype due to lack of the targetable receptors ER, PR and HER2 (Triple Negative Breast Cancer). Systemic chemotherapy is the only available option and unfortunately, only 20% of tumors respond to the standard chemotherapy.

Triple negative breast cancer (TNBC) that includes the basal subtype is highly heterogeneous than was thought before and gene expression analysis identified six more of its subtypes, each with unique molecular markers. The six subtypes include basal-like (BL1 and BL2), immunomodulatory, mesenchymal (M), mesenchymal stem-like (MSL) and luminal androgen

receptor (LAR) subtype (Mayer 2014). Owing to this heterogeneity, no single targeted therapy has been approved for the treatment of TNBC.

Luminal A subtype breast tumors are most common (70%) while the TNBC subtype tumors account for 15%. The TNBC and the HER2 overexpressing tumors have the worst prognosis. The metastatic disease is the cause for the estimated 90% deaths due to breast cancer and is not curable. De novo and acquired resistance are the major challenges and no targeted therapies have been approved for the most aggressive TNBCs.

Single targeted therapies including platinum agents, PARP inhibitors, PI3K and MEK inhibitors did not bear favorable outcomes for TNBC treatment. No single targeted therapy has been approved for TNBC and the cytotoxic chemotherapy remains the standard treatment. Apart from the standard chemotherapy, immunotherapy targeting CTLA-4 and PD-L1 are investigated for treatment of TNBC. Combination of two or more agents may make a more rational approach to treat TNBC. The most commonly used combination chemotherapies show only 7-11% improvement in survival for younger women and just 2-3% for women above 50 years.

30% of all breast cancer tumor subtypes carry mutations in *TP53*. The distribution of *TP53* mutations in the subtypes include 26% in luminal tumors (17% luminal A, 41% luminal B), 88% in basal-like or TNBC carcinomas, and almost 50% in HER2+ tumors. The high rate of *TP53* mutations in TNBC promotes survival and proliferation of the cancer cells, contributes to the heterogeneity and resistance to chemotherapy. The current therapeutic strategies aim at restoring the wildtype activities of the p53 tumor suppressor assuming that the different p53 missense mutants are functionally equal. Thus, so far, no drug has been effective in abrogating the oncogenic effects of mutant p53. The non-specific chemotherapy benefits only a small fraction of patients as we do not know how the heterogeneity of each tumor affects response to therapy or the recurrence. The TNBC tumors contain more genetic mutations compared to luminal and HER2+ subtype due to high mutation rate and identification of the minor subclones

(infrequent but biologically important somatic mutations) and designing targeted therapies against them could be effective in treating the TNBC patients.

Overarching goal

The focus of the research is to gain a mechanistic understanding of the phenotypic heterogeneity observed in aggressive triple-negative breast cancer (TNBC) harboring different *TP53* mutations and numerous other somatic mutations to discover novel therapeutic targets for personalized medicine. The phenotypic and transcriptomics data on 10 normal breast epithelial cell lines each expressing a distinct missense mutant p53 protein and systems-level analysis through ChIP Seq, RNA seq and CRISPR-based screening would enable identification of possible novel molecular targets for personalized treatment of TNBC patients with specific *TP53* mutations. Such a broad-spectrum molecular profiling and functional genomics screening could efficiently select and narrow down the targetable co-existing mutations of mutant p53 and their downstream targets and pathways as key candidates for personalized targeted therapies.

This chapter provides a comprehensive overview on the tumor suppressor gene *TP53* in humans. The first section covers the structure and function of the tumor suppressor protein. The initial discovery of *TP53* is a fascinating story because of the prevailing view at the time that *TP53* acts as an oncogene instead of a tumor suppressor gene. The following sections describe the mutation repertoire of *TP53* in cancer, its contribution to cancer initiation and progression in general and specific contribution towards breast cancer progression. *TP53* remains the most frequently mutated gene in all cancer types and exists in a sea of other somatic mutations. Mutant *TP53* is not effective in initiating cancer alone and interacts with the other co-mutations or the co-drivers. The last section of the chapter describes clonal evolution of cancer tumors, the interacting somatic mutations and the tumor heterogeneity they generate. Without a doubt, the

TP53 gene is the most extensively studied tumor suppressor gene. Yet, in the 40 years since its discovery, we still do not fully understand p53's functional complexity.

TP53: Tumor suppressor frequently mutated in all cancer types

Function and regulation of the p53 tumor suppressor protein

TP53 tumor suppressor gene controls multiple cellular programs that work towards suppressing cancer and that explains why this gene is mutated in multiple cancer types. A range of stresses activate p53, including DNA damage, hyperproliferative signals, nutrient deprivation, hypoxia, oxidative and replicative stress. p53 mediates tumor suppression by transient cell-cycle arrest, senescence and apoptosis that have been termed as the canonical p53 functions. But recently the focus has shifted to the non-canonical functions of p53 that include regulation of cell metabolism, stem cell maintenance, invasion and metastasis control to mediate tumor suppression (Fig 1.1). The loss of p53 influences cell motility and contributes to invasive and metastatic potential (Kaiser 2018). Wild-type p53 prevents epithelial to mesenchymal transition (EMT) by degrading Slug and enhancing E-cadherin expression strengthening cell-cell junction. Loss of p53 activity is accompanied by increased activity of RhoA/ROCK promoting cell motility and invasion during tumor progression. Mutant p53 suppresses E-cadherin while activating Slug or Twist inducing EMT (Muller 2011).

The p53 is widely known as the 'guardian of the genome' by mediating tumor suppression via the classical function of inducing transient G1 cell-cycle arrest, DNA repair, apoptosis and senescence restricting the propagation of DNA damaged cells (Beiging 2014). Both the canonical and the non-canonical functions of p53 demonstrate the diverse phenotypes the protein controls and why *TP53* is mutated in a wide variety of cancer types (Fig 1.2).

The action of tumor suppressor p53 is context dependent. The determinants of this context dependency include the cell type, genetic background, cell microenvironment and the nature of the stress that activates p53. The mechanism for context dependence may operate depending on amount of p53 protein induced, post translational modifications of p53 and other parallel signaling networks. Both canonical and non-canonical responses of p53 integrate to suppress tumor formation.

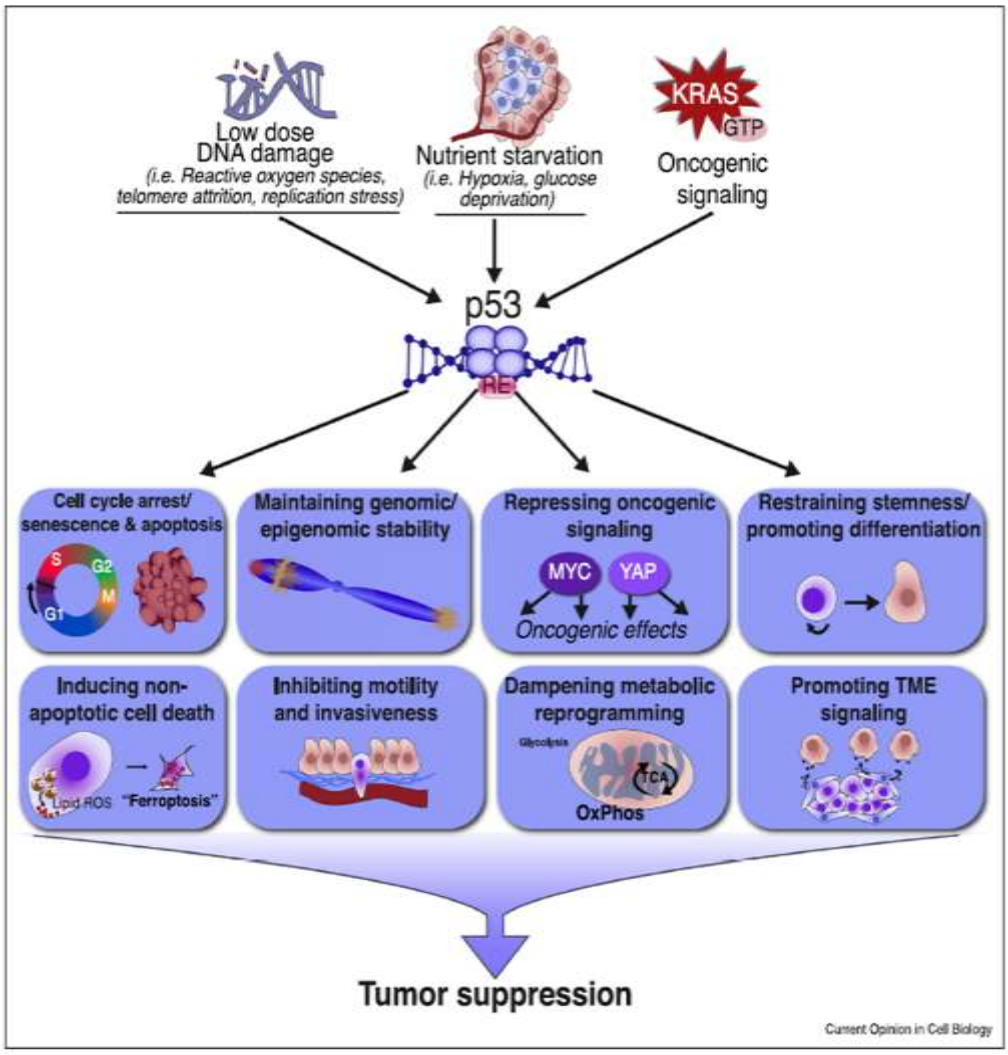


Figure 1.1. **Signaling pathways regulated by p53 in tumor suppression.** Low dose DNA damage, nutrient starvation and oncogene activation are some of the stresses encountered in a tumor that activate p53. In response, p53 binds to the relevant response elements and regulates the gene expression coordinating different cellular processes contributing to tumor suppression. Image adapted from (Mello 2018).

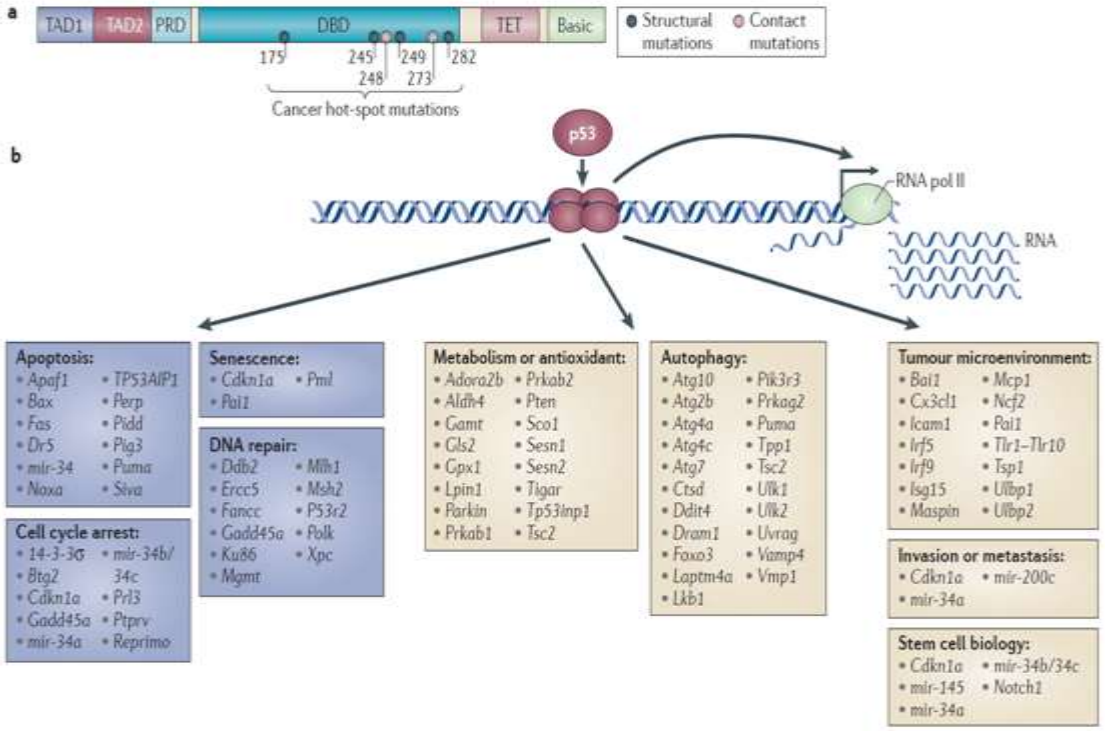


Figure 1.2. **The p53 target genes.** A. The domains of the p53 protein. B. The key p53 target genes for tumor suppressive activity of p53. The canonical functions are in purple and non-canonical in beige. Image adapted from (Beiging 2014).

TP53 is a transcription factor and binds to its DNA response elements in a sequence specific manner via the DNA binding domain and regulates transcription by the two N-terminus transactivation domains (TADs) and the regulatory C-terminal domain (CTD). Each TAD transactivates different target genes and effector pathways and makes p53 a robust transcription factor. The TADs are known to interact with diverse transcriptional cofactors and may contribute to p53 function in a context dependent manner. CTD is a positive regulator of p53 function and needed for transactivation of the target genes. The CTD is intrinsically disordered informing that it could bind to several different p53 transcription co-factors and other binding partners depending on the context thus, affecting the p53 function. The CTD also serves as a site for many post translational modifications (PTMs) (Sullivan 2018).

Under cellular stress p53 protein is stabilized and the transcriptional output depends on the strength of p53 response element, PTMs, the available binding partners and epigenetic landscape of target gene promoters (Beckerman 2010). Hundreds of PTMs have been reported for p53 and are crucial for regulating p53 level and activity (Fig 1.3). Phosphorylation of p53 is essential for p53 stabilization. Polyubiquitination of p53 by Mdm2 controls p53 protein levels and phosphorylation of p53 at the N-terminal weakens this interaction and in turn stabilizes p53 in response to cellular stress signals. Specific PTMs occur in response to different stress signals and direct different p53 responses and cell fate. For example, phosphorylation at Ser15 by ATM and ATR kinases is required for cell cycle arrest whereas phosphorylation at Ser46 by HIPK2 regulates apoptosis. Site-specific acetylation at CTD by p300 or PCAF regulate apoptosis and cell cycle arrest. Methylation of p53 at different lysine residues can activate or repress gene transactivation. Mutant p53 proteins show intense phosphorylation and acetylation at the sites known to stabilize wildtype p53 and results in accumulation of mutant p53 in the nucleus (Bode 2004). Sumoylation and neddylation are some of the other PTMs undergone by p53 that contribute to p53 promoter specificity.

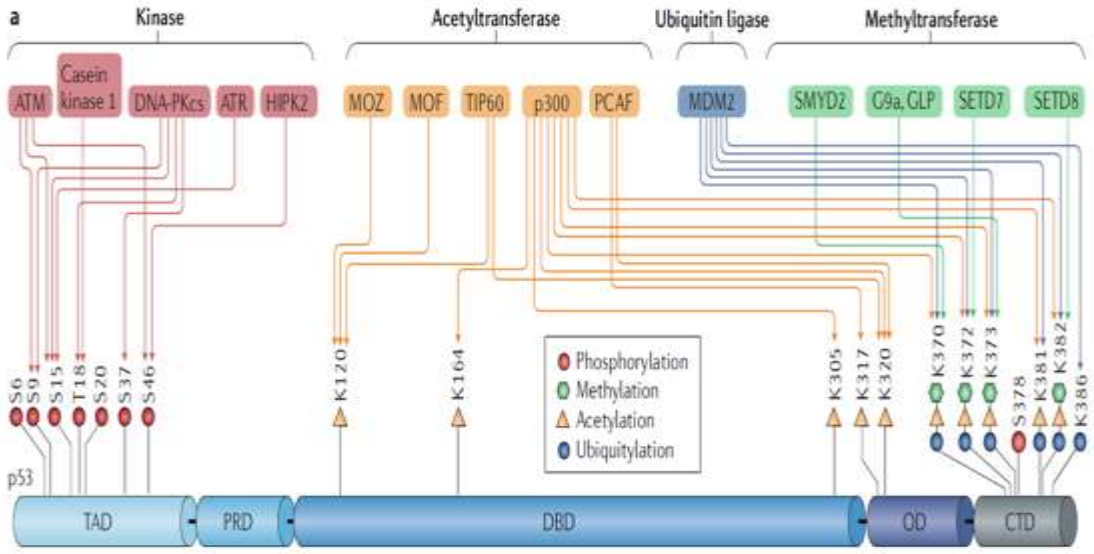


Figure 1.3. **Different PTMs of p53 and the enzymes that catalyze them.** Phosphorylations occur at N-terminus. Acetylation in DNA binding domain (DBD), oligomerization domain (OD) and carboxy terminal domain (CTD). Methylation and ubiquitylation occur at the CTD. Image adapted from (Hafner 2019).

Transcription co-factor recruitment by p53 affects the activation of specific gene program like p53 interacts with acetyltransferase p300 and activates the target gene *CDKN1A*. The p53 level in the cells change over time and so the gene expression dynamics of the p53 target genes showing different timing, level and pattern of induction. Overall, p53 PTMs and cofactors regulate stability, strength of DNA binding and target gene selection (Hafner 2019). With p53 binding to its response element within a promoter recruits chromatin remodeling factors, histone transacetylases and methyltransferases that result in histone modification and promoter opening. The approximate number of p53 target genes is 300 and the functional p53 binding sites are independent of cell type and treatment (Fischer 2017). Recent studies define p53 as only a transcription activator and not a direct transcription repressor.

The discovery of p53 and its establishment as a tumor suppressor gene

In the period between 1950 and 1970 much attention in research was grabbed by viruses and oncogenes. By the 1970s the researchers found that many viruses carried oncogenes. The Bishop and Varmus labs discovered that the RNA tumor viruses captured a cellular gene and reintroduced its modified version when infecting a cell whose overexpression gradually caused the transformation of the cell. By then it was established that oncogenes caused cancer in animals and that the DNA tumor viruses acted the same way as the RNA tumor viruses. The oncogenes from DNA tumor viruses were different from the RNA tumor viruses and believed to indirectly induce excessive production of oncoproteins (Levine 2009).

In the initial years of discovery of *TP53*, now known as a major tumor suppressor gene, was instead accepted as an oncogene. In the year 1979, researchers found that the small DNA tumor virus, the simian virus 40 (SV40) contained two antigens small t and the large T antigens and the large T antigen initiated and maintained the transformation of the infected cells. While understanding the mechanism of cellular transformation by SV40, T antigen formed a complex with a specific cell coded protein in SV40 transformed mouse cells with a molecular weight of 53

kDa. The labs of Lane and Crawford inferred that either this 53 kDa protein had the same antigenic determinants as the T antigen or it forms a complex with the T antigen (Lane 1979). Similar results from another study showed a 54 kDa protein immunoprecipitated from a SV40 transformed mouse cell line and uninfected embryonal carcinoma cells with SV40 anti T sera. The immunoprecipitated proteins had identical peptide maps but completely different from peptide map of large T antigen. The researchers concluded that the transformation of the mouse cells with the SV40 virus stimulated the synthesis or stability of the 54 kDa protein, which associated with the large T antigen in the infected cells (Linzer 1979). In yet another study, the sera from BALB/c mice bearing tumors from injection of Abelson murine leukemia virus (A-MuLV) transformed cells reproducibly immunoprecipitated a 50 kDa protein from A-MuLV transformed BALB/c cells. The 50 kDa protein remained at very low levels in normal cells or cells not transformed by A-MuLV (Rotter 1980). As high levels of this protein presented in virus transformed cells (SV40 or A-MuLV) and hardly detectable in non-transformed cells, it was convincing that this protein is a tumor antigen and thus, a cancer-causing oncogene. This protein became designated as the 'Tumor protein 53 (*TP53*) or simply p53'.

Several observations made p53 protein look like a cellular oncogene. Normal mouse 3T3 cells had low levels of p53 protein which escalated to 25-50-fold with transformation by SV40. But a temperature sensitive mutant of SV40 large T antigen could not maintain a high level of the p53 protein in the virus infected cells (Linzer 1979). The E1b protein from another DNA virus, adenovirus, associated with the 54 kDa protein (p53) in adenovirus transformed mouse cells (Sarnow 1982). The p53 protein was also found at high concentrations in transformed cell lines and in induced and spontaneously occurring mouse tumors making it a biochemical diagnostic marker for primary tumors (Rotter 1983).

After the discovery of *TP53,* the study of its 'oncogenic properties' required cloning of the gene. As transformed cells contained abundant p53, it reasoned that the cells would have more p53 mRNA and were used to prepare the first cloned cDNA instead of normal cells. The p53 protein

from the cloned gene cooperated with *HRAS* oncogene to transform the normal embryonic cells and these cell lines formed tumors in animals (Eliyahu 1984). In another study, the co-transfection of murine cellular p53 gene with *Ras* gene into primary rat embryo fibroblasts converted the normal cells into tumorigenic fibroblasts (Parada 1984). When a functional p53 gene was transfected into L12 cells (Ab MuLV transformed cells lacking p53), p53 expressing L12 cells caused lethal tumors in mice (Wolf 1984). Gathering the results, overexpression of p53 in transformed cells presented *TP53* as an oncogene that drove cancer progression.

But studies since the 1980s completely changed the way researchers had described the *TP53* gene. There were increasing evidences that *TP53* was a tumor suppressor and not an oncogene. These studies were an eye opener and showed *TP53* into new light as a potent tumor suppressor gene whose wildtype version was frequently deleted in transformed cells and tumors. A study in the mid-1980s reported that Abelson murine leukemia virus transformed L12 cells lacked expression of cellular p53. The p53 genomic DNA sequence was interrupted by integration of a viral DNA segment (Wolf 1984). This provided one of the first evidences that the p53 protein could be a tumor suppressor protein rather than an oncogene. Analysis of the genomic DNA in the HL-60 cells (human tumor cell line) lacked the p53 gene due to deletion which contrasted with the previous studies where the p53 tumor antigen was found to be over-produced in the transformed cells (Wolf 1985). The studies implicated that the loss of functional p53 promoted cancer and the full-length protein is required to prevent cancer. When the DNA sequences of the various p53 clones were compared, none of the sequences were identical suggesting that different clones contained mutations in the *TP53* gene. The wildtype p53 cDNA failed to transform the primary rat fibroblasts while mutated p53 resulted in transformation (Finlay 1988). In the tumors of colorectal carcinoma, p53 was lost due to allelic deletion in chromosome 17. The remaining allele of p53 was mutated where a single amino acid was substituted with another. Moreover, the expression of wildtype p53 also inhibited the transformation of primary rat embryo fibroblasts by E1A and Ras (Finlay 1989). By the end of 1980s, substantial evidences were found to highlight *TP53* as a tumor suppressor gene and now called the '*Tumor suppressor Protein 53'*.

Somatic mutations in *TP53* are abundant in cancer cell lines and the tumors. Germline mutations in *TP53* were discovered in the Li-Fraumeni patients that predisposed patients to increased cancer susceptibility (Srivastava 1990). In mice models, homozygous deletion of the p53 alleles did not affect development but made the mice prone to spontaneous neoplasms by 6 months of age. The absence of *TP53* predisposing mice and Li-Fraumeni syndrome patients to neoplastic disease undoubtedly presented *TP53* as a tumor suppressor gene (Donehower 1992). Finally, by the year 1989, 10 years since the discovery of p53, it was defined as a tumor suppressor gene and shown to be either mutated or lost in the human cancer tumors. The researchers did not have the slightest idea how versatile this protein could be. While the wildtype p53 is a potent tumor suppressor, the mutant p53 exerts a dominant negative effect on wildtype p53 (if present) as well as can have gain-of-function (GOF) activities like oncogenes in the absence of wildtype p53 alleles.

The structure of the p53 protein

The *TP53* tumor suppressor is a transcription factor and shows sequence specific DNA binding. A good understanding of the structure facilitates a deeper understanding of the structure-function relationship and how the mutations at specific locations affect the conformation of the protein and the resultant function of the mutant form.

The p53 transcription factor is functional as a homo-tetramer and the N-terminus of a monomer contains an intrinsically disordered trans-activation domain (TAD) and a proline rich region. The central domain is folded and is the DNA binding core domain that facilitates sequence specific DNA binding (Fig 1.4). A flexible linker connects this domain with a tetramerization domain that enables the homo-tetramer formation. At the C-terminus of the monomer protein is a regulatory domain. This is an unfolded region rich in basic amino acids (lysine) and non-specifically binds DNA (Joerger 2008).

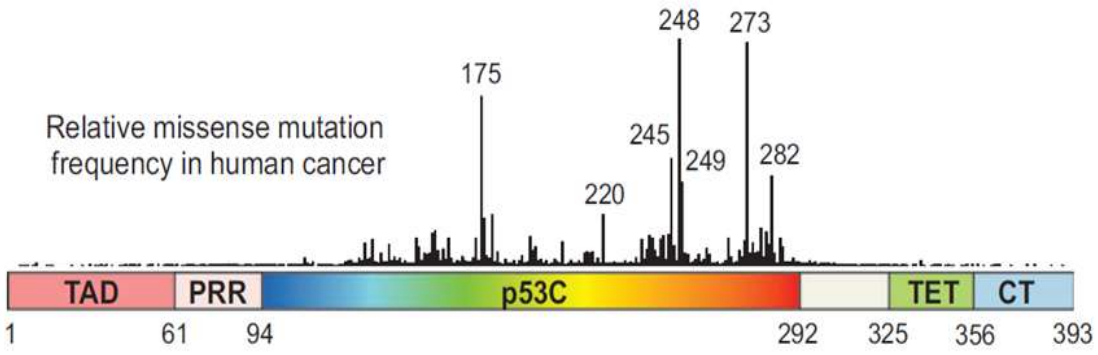


Figure 1.4. **Domains of the full length p53 protein.** N-terminal transactivation domain (TAD), proline-rich region (PRR), central core DNA binding domain (p53C), tetramerization domain (TET) and the C-terminus (CT). The rainbow-colored DNA binding domain is where most cancer associated mutations occur. The bars represent the relative frequency of missense mutation at the residue. Image from (Joerger 2008).

The N-terminal region of the monomer is intrinsically disordered. The TAD in this region is a binding site for several interacting proteins of p53. For example, the transcriptional co-activators p300/CBP and the negative regulators MDM2/MDM4. This intrinsic disorder in TAD is often seen in proteins that form a part of a signaling network and promotes binding to diverse other proteins with high specificity. The residues undergo disorder-to-order transition on binding the interacting proteins or nucleic acids. The proline rich region contains five PXXP motifs that are known to interact with proteins that contain Src homology 3 (SH3) domains. It is likely that these poly-prolines provide rigidity to the formed helices in the TAD regions and project the TAD away from the central core domain (Wells 2008).

The central DNA-binding core domain is where the missense mutations in p53 concentrate. It contains a β-sandwich that forms the primary scaffold for DNA-binding surface. The surface has two structural motifs i) loop-sheet-helix motif that bind the DNA major groove. It consists of loop L1, β-strands and C-terminal helix. ii) L2 and L3, two large loops contact the DNA minor groove and are stabilized by a zinc ion. The ion is tetrahedrally coordinated by histidine at position 179

and cysteines at 176, 238 and 242. The loss of zinc results in instability and structural distortion of the domain affecting DNA binding specificity of the p53 monomer (Fig 1.5).

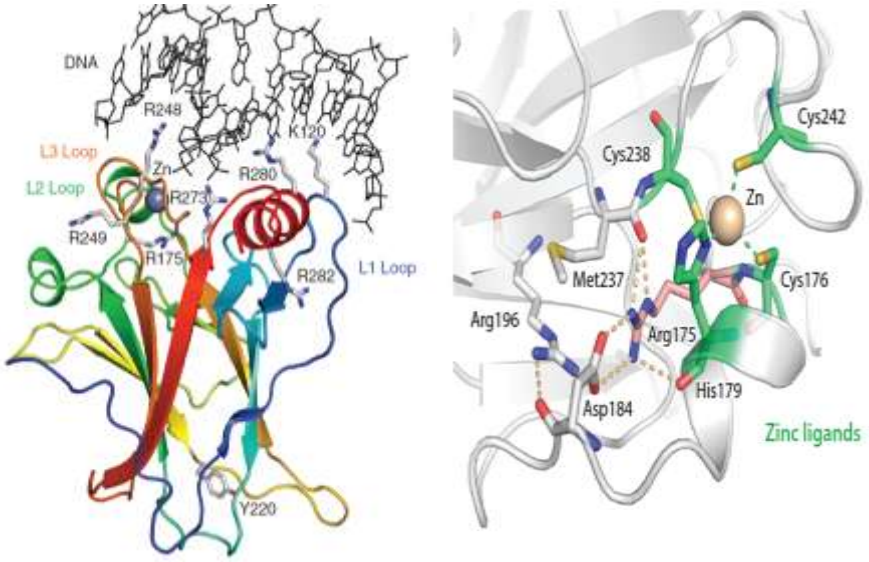


Figure 1.5. **Ribbon representation of the DNA binding domain of p53.** Amino terminus in blue and carboxyl terminus in red. The essential DNA contact points are shown as grey sticks. On the right, p53 zinc coordination sphere and the amino acids involved are in green. Images adapted from (Joerger 2016, Joerger 2010).

The p53 tetramer has sequence specific DNA binding capability and the target binding sites consist of two decameric motifs (half-sites) of the general form RRRCWWGYYY (R=A, G; W=A, T; Y=C, T), separated by 0–13 base pairs. But most p53 binding sites have consecutive half sites. For sequence specific DNA binding two core domains bind to a half site forming a dimer. The L3 loop contacts the DNA minor groove through Arg248 and Arg273 contacts the DNA phosphate backbone. The functional p53 is described as a dimer of dimers and two dimers binding to the half sites form a tetramer which is stabilized by base stacking and protein-protein interaction (Joerger 2010) (Fig 1.6). As the DNA binding domain determines the sequence specificity of p53, missense mutations (replacement of one amino acid with another in the protein) frequently occur in this domain and contain the mutation hotspots that is described in the next section.

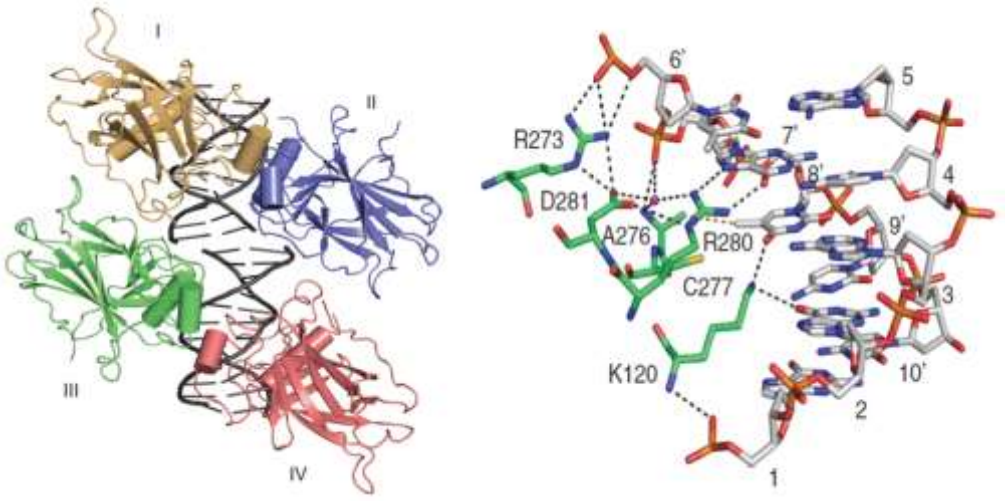


Figure 1.6. **Core domain tetramer bound to DNA in a sequence specific manner**. A p53 dimer binds to the palindromic half site on the DNA. Right, DNA contact residues in the major groove interaction. Image adapted from (Joerger 2010).

The full length p53 monomers form the functional tetramers through the tetramerization domain in the C-terminal region of p53. These tetramers are reversible in nature. The tetramerization domain in each monomer has a short β-strand and a α-helix linked by a sharp turn. The tetramer is formed from the dimerization of the dimers (Fig 1.7). Hydrophobic interactions contribute towards the stabilization of the p53 tetramer.

The C-terminal regulatory domain is also intrinsically disordered and subjected to PTM (described under function and regulation of p53). Acetylation of lysines at C-terminal increases sequence specific DNA binding and may modulate p53 transcription activity. This domain interacts with different regulatory proteins and may have a role in p53 activation, degradation and localization. On binding other proteins or non-specific DNA, a disorder-to-order transition occurs in the C-terminal domain which adopts different conformations based on structural context and PTM patterns and hence, referred to as a 'chameleon sequence' (Joerger 2010). Recently, it was

confirmed that the lysine rich C-terminal domain (CTD) governs the binding of p53 in a DNA sequence dependent manner and stabilizes the p53-DNA complexes (Laptenko 2015).

The knowledge of the quaternary structure of p53 is necessary to understand its function. In the open conformation, the core domains are accessible to DNA. On binding DNA, p53 wraps around it and the overall structure becomes less flexible. Four p53 core domains bind to the consensus sequence forming a tetramer. In both free and DNA bound p53, the N-termini remain extended and suggest their

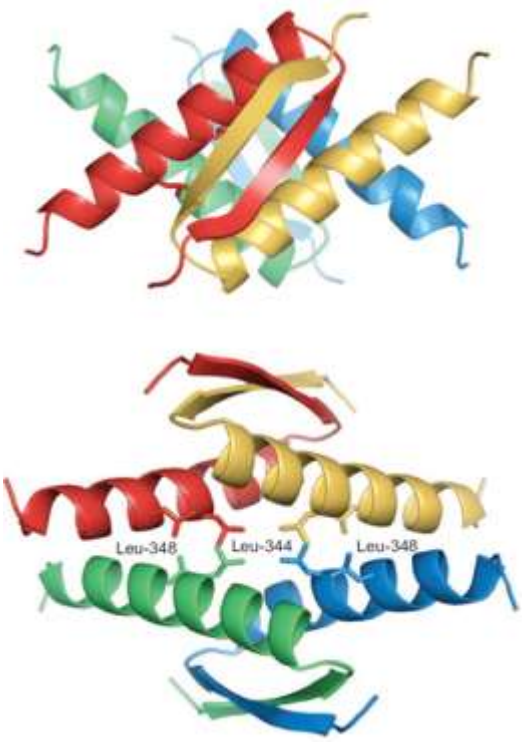


Figure 1.7. **Dimerization of the p53 dimers forming the functional tetramer.** Shown in two separate orientations. Image taken from (Joerger 2008).

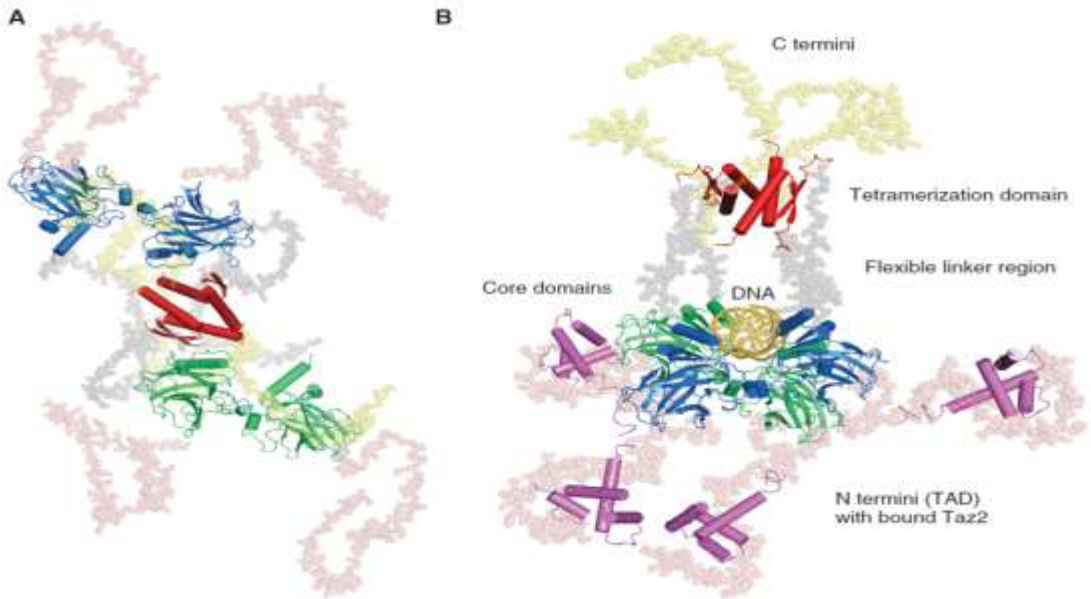


Figure 1.8. **Full length p53 in unbound state** in A. and B. with bound DNA showing the ternary complex (homo-tetramer). DNA binding domains (blue, green), tetramerization domain (red) and TAD (purple). Image from (Joerger 2010).

involvement with different proteins and post-translational modifications (Fig 1.8). Overall, the intrinsically disordered N and C-terminal domains provide p53 structural and thus, functional plasticity needed to bind to different protein partners.

TP53 mutations in human cancer

Cancer associated somatic mutations in the TP53 gene

The TP53 database of the International Agency for Research on Cancer (IARC TP53 database) collects information on *TP53* mutations in all types of cancer from peer reviewed literature and available databases. The database is updated regularly, and the latest version is R19. The database includes information about *TP53* somatic mutations in sporadic cancers, germ-line mutation in familial cancers, *TP53* polymorphism, functional assessment of the mutant protein, mutation in cell lines and *TP53* mouse models. The database has been helpful in analyzing tumor specific *TP53* mutation patterns, genotype and phenotype relationship and clinical impact of mutant p53 proteins.

The database has reported that majority of the mutations in the sporadic cancers are missense mutations caused by single nucleotide substitutions in *TP53* that account for about 75% in total (Fig 1.9). The remaining 25% of the mutations include truncation (nonsense, frameshift and deletion) and silent mutations. These missense mutant p53 proteins are full length proteins with a single change in the amino acid sequence characteristic of activated oncogenes. The highest frequency of missense mutations in *TP53* has been observed in ovarian carcinomas in contrast to cervical carcinomas with much less frequent mutations in *TP53*. Somatic mutations in *TP53* are also found in colorectal, lung, esophagus, liver, breast, brain, bladder, leukemia and skin cancers (Petitjean 2007). These missense mutations primarily occur in the evolutionarily conserved codons of *TP53,* but the mutational spectrum differs among different cancer types. The mutation

spectrum differs in the position of the hotspots (frequently mutated nucleotide), frequency of transitions (purine/pyrimidine substituted by purine/pyrimidine) and transversions (purine replaced by pyrimidine or vice versa). For example, colon tumor mutations are mainly caused by base transitions at CpG dinucleotides. In comparison, G to T transversions were much higher in the breast tumors (Hollstein 1991).

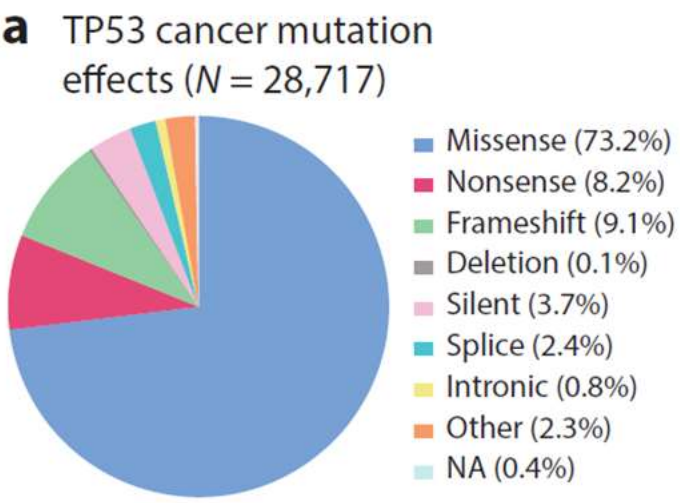


Figure 1.9. ***TP53* mutations in cancer.** The different cancer associated somatic mutations in *TP53* based on IARC TP53 mutation database (Release R17, November 2013, *N*=28,717, **http://p53.iarc.fr/).** Taken from (Joerger 2016).

The frequent somatic missense mutations found in human cancers are the most commonly observed genetic alterations clustered within the evolutionarily conserved DNA binding domain of the p53 protein and uncommon in the amino and carboxy terminus of the protein (Fig 1.10). The non-sense, silent and frameshift mutations predominate in the N and C-terminal regions. Among the missense mutations occurring in the DNA binding domain there are a few frequent mutations that account for nearly 30% of all missense mutations in cancers and have been addressed as the 'hotspot' mutations and underline the importance of these sites for p53-DNA complex formation. The frequently mutated amino acid residues are located near the protein DNA interface. They occur in the L3 loop (minor groove contact), the loop-sheet-helix motif (major

groove contact) and the L2 loop (interacts with L3 loop). The six major hotspots include Arg248, Arg273, Arg175, Gly245, Arg249 and Arg282 (Fig 1.11).

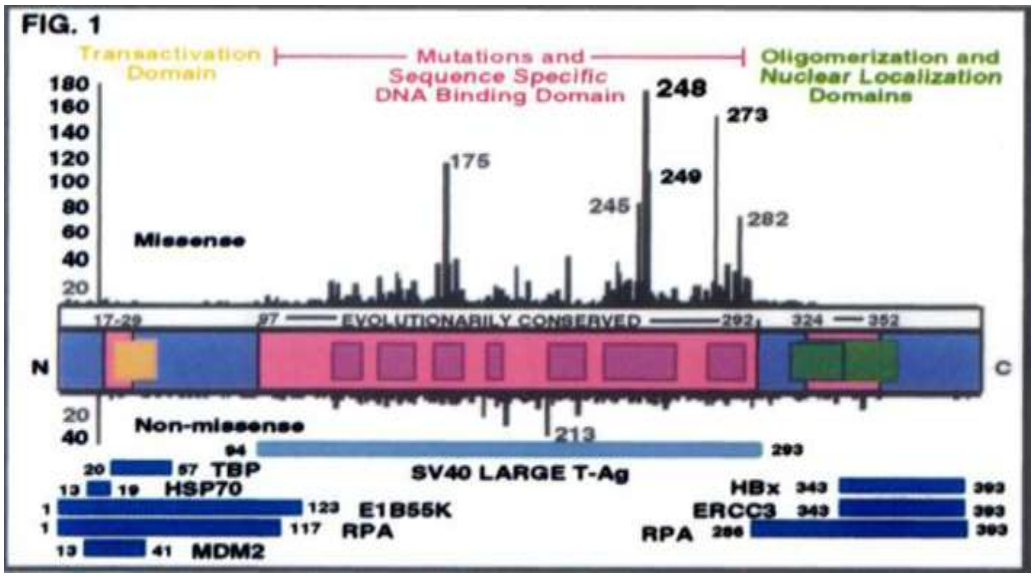


Figure 1.10. **Evolutionarily conserved domains (purple boxes) in the DNA binding domain of p53 protein (pink).** Vertical lines above denote missense mutations (mainly in the conserved regions) with the height corresponding to its frequency and below denote non-missense mutations (occurring throughout the protein). Adapted from (Greenblatt 1994).

Arg248 in the L3 loop contacts the minor groove of DNA and Arg273 from the loop-sheet-helix motif contacts the phosphate backbone of DNA in the major groove. The rest of the hotspot residues stabilize the structure of DNA binding surface of p53 transcription factor. Arg175 is in the L2 loop and stabilizes the zinc ion binding that bridges the L2 and L3 loops. Arg249 is in the L3 loop right next to Arg248 that contacts the DNA minor groove. Arg282 resides in the loop-sheet-helix motif and Gly245 in the L3 loop is the only non-arginine residue important for the maintenance of the L3 loop conformation. These hotspot mutants can be classified into two groups- mutations in the residues that directly contact the DNA and cause loss of critical DNA contacts (DNA contact mutants) and mutation of residues that stabilize the folding of the core domain and may cause structural defects (structural mutants). Mutation at Arg175 is an example

of structural mutant. Other examples for structural mutants include Arg249, Gly245 and Arg282. The mutations at Arg273 and Arg248 are referred to as the DNA contact mutants that abrogate important DNA contacts at the p53 protein-DNA interface but do not have any significant effect on the structure of DNA binding core region of p53 but influence the DNA binding affinity of p53. Thus, DNA contact mutants inactivate p53 by losing critical DNA contacts and the structural mutants disrupt the structural integrity of p53 core domain and inactivate it (Cho 1994). The five hotspot codons (175, 245, 248, 273 and 282) contain CpG dinucleotides and the single base substitutions result from G:C to A:T transitions (deamination of 5-methylcytosine) at these sites. Deletions and insertions may result from DNA polymerase infidelity during replication or faulty repair of the carcinogen-DNA adducts (Greenblatt 1994).

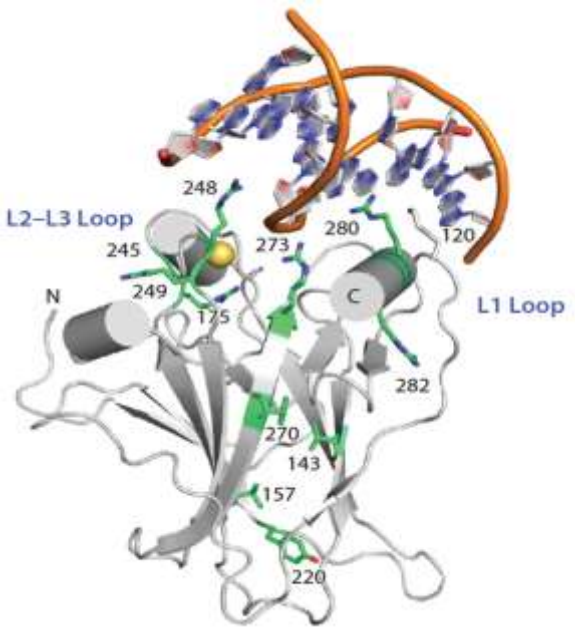


Figure 1.11. **p53 DNA binding domain bound to a DNA sequence.** The hopspot mutation sites include codons 175, 248, 273, 245, 249 and 282. Figure adapted from (Joerger 2016).

The most common genetic aberrations in other tumor suppressor genes such as *RB* and *APC* are deletions or non-sense mutations in both the alleles that knock-out proteins and their loss-of-function leads to tumorigenesis. In stark contrast, the *TP53* tumor suppressor undergoes not only deletion but also frequent missense mutations generating a full-length mutant protein like tumor causing oncogenes, with eventual loss or mutation of the second wildtype allele.

Germline mutations in TP53 gene and Li-Fraumeni syndrome

The Li-Fraumeni syndrome (LFS) is an autosomal dominant disease where patients are pre-disposed to cancer with early onset of tumors (strikingly lower than sporadic cancers) including sarcomas, brain tumor, breast cancer and adrenocortical carcinoma. Germline mutations in *TP53* form the primary molecular basis for LFS development. 74% of these mutations are missense mutations that predominantly occur in the highly conserved DNA binding domain of p53 including the hotspot codons (Guha 2017). The mutations also include non-sense and splice mutations distributed across the *TP53* gene. Mutation in one allele is followed by loss of heterozygosity (LOH) in the other either by mutation or deletion. Thus, *TP53* qualifies as a tumor suppressor gene whose inactivation leads to tumor development as per 'Knudson two hit hypothesis'. This was proposed for genetic model of retinoblastoma development where inheritance of one mutated tumor suppressor RB allele was the first hit and mutation or loss of second RB allele as the second hit promoting retinoblastoma development (Rivlin 2011).

LFS patients have increased risk of developing multiple primary tumors and the risk of developing cancer in patients with *TP53* mutations is approximately 73% in males and abnormally high in females being 93% where they develop breast cancer at a very young age. The clinical presentation of LFS is very heterogeneous as the age of tumor onset and the subtype vary and suggests the involvement of other genetic factors apart from germline mutation in *TP53* for example, mutation in *BRCA1*.

Trp53 germline mutant heterozygous mice present a close genetic model of LFS as patients are heterozygous for *TP53* and develop metastatic tumors providing evidence for dominant negative and eventual gain-of-function phenotypes of mutant *TP53*. In the identical genetic background two different missense mutations, R172H and R270H (human R175H and R273H), gave rise to different tumor spectra observed in humans with LFS (Olive 2004). The loss of the wildtype *TP53* allele was significantly lower in the tumors carrying missense mutation in the DBD of *TP53* (Birch

1998). These observations ascertain that some mutants exhibit dominant negative effect over wildtype allele or gain-of-function properties when wildtype allele is lost. The mean age of onset of breast cancer is much lower in LFS families carrying a germline *TP53* mutation than LFS families without *TP53* mutation. The Li-Fraumeni syndrome where germline mutation in *TP53* predisposes individuals to cancer underlines the importance of *TP53* tumor suppressor in restricting cancer development.

Somatic mutation burden in cancer tumors

The somatic cells accumulate spontaneously occurring mutations over time. These mutations arise due to errors in DNA replication or when a damage in the DNA is repaired incorrectly or kept unrepaired. A small fraction of these spontaneous mutations provides selective advantage in terms of growth and survival of cells and are termed the 'driver' mutations that remain under positive selection. 1-2 % of the protein coding genes in human genome are recurrently mutated in cancers and are the drivers that drive the process of oncogenesis. Most mutation variants do not provide any selective biological advantage and are termed the 'passenger' mutations. There is a small fraction of advantageous infrequent mutations that co-appear with the major drivers and contribute towards tumor development. They can be addressed as the 'co-drivers' working as an accomplice to the drivers. The targeting of the pathways involving the co-drivers in addition to the driver pathway may enhance the effect of targeted combination therapies for cancer treatment.

The advancement in the genome sequencing methods over the years has revealed the repertoire of somatic mutations in different cancers. The results highlight the mutational processes and the 'cancer genes' (Martincorena 2015). Most of the cancer tumors carry 1000 to 20,000 somatic point mutations and hundreds of insertions, deletions and gene rearrangements, DNA copy number increase as well as reductions (Fig 1.12). The cancer cells also acquire epigenetic changes that manifest as changes in DNA methylation and chromatin structure. The overall somatic mutation prevalence differs with the cancer types. The lung carcinomas followed by

gastric, ovarian and colorectal cancers show highest prevalence of somatic mutations and may carry more than 100,000 point mutations. On the other hand, hematological cancers and most breast cancers show a comparatively lower prevalence of somatic mutations (Greenman 2007). High cell turn-over and exposure of the surface epithelia to exogenous mutagens as in colorectal, lung and gastric cells leads to high somatic mutation rates.

The analysis of 9,423 tumor exomes using 26 computational tools has identified 299 driver genes and more than 3000 putative missense driver mutations (Bailey 2018). To date, this study has been the most comprehensive discovery of cancer genes and mutations. The mutation spectrum and cancer driver genes vary among different cancer types. Some driver genes are specifically associated with a single cancer while some have driver roles in two or more cancer types. As an example, *TP53* gene acts as a driver in 27 cancer types and *PIK3CA, KRAS* and *PTEN* are

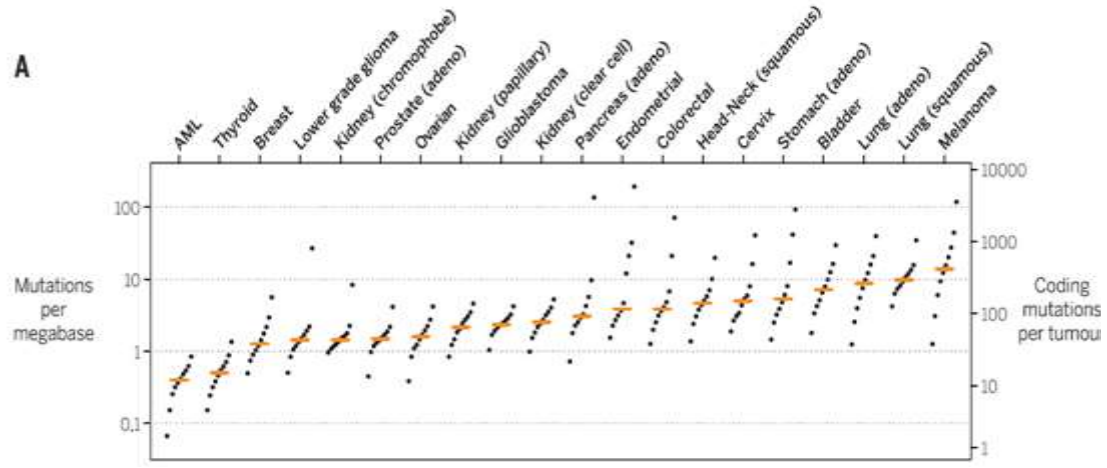


Figure 1.12. **Somatic mutation burden (base substitutions and small indels) in 20 different types of cancer.** The orange bar is the median burden of all samples. Image adapted from (Martincorena 2015).

associated with 15 or more cancer types. As evident from the sequencing of tumors, cancer is a disease of numerous mutations that are required for turning the benign tumors into the malignant form. Apart from the driver mutations, additional less prevalent mutations 'the co-drivers' as introduced before, exist that do not clonally dominate the tumor but make reservoir of genetically

heterogeneous cells that may contribute to metastasis or drug resistance and to tumor heterogeneity. The immediate challenge lies in identifying these less prevalent drivers, the co-drivers, in the sea of passenger mutations.

TP53 mutations in breast cancer

Germ-line mutation in *TP53* is found in patients of Li-Fraumeni syndrome with increased risk of breast cancer and this established the involvement of *TP53* mutations in breast cancer development. Somatic mutations in *TP53* are also known to occur in Ductal Carcinoma In-Situ (DCIS), preceding stages of invasive breast cancer and the frequency of p53 mutations increase with the grade of DCIS tumors. These imply that *TP53* mutations occur early in the process of breast carcinogenesis (Boressen-Dale 2003) and has been designated as one of the driver mutations. The frequency of *TP53* mutations are found to be much higher in node-positive, large tumors, tumors from advanced disease and recurrent tumors than primary tumors.

Breast cancer is a highly heterogeneous disease with tumors of characteristic molecular features, prognosis and therapy responses. Gene expression patterns obtained from cDNA microarray on breast tumors and normal breast tissues classified the tumors into basal type, ERBB2 or HER2 overexpressing, normal breast like group and luminal type (with subgroups luminal A and luminal B) each with distinctive expression profiles (Table 1.1). Poor prognosis is associated with the basal-like subtype and happens to be the most aggressive of all subtypes (Sorlie 2001). Immunohistochemically, luminal subtype is estrogen receptor (ER+), progesterone receptor (PR+) and HER2 receptor positive (HER2+). HER2 overexpressing subtype has amplification of HER2 receptor and could be ER+ or ER-. The hormone receptor positive subtypes are treated with hormone therapy targeting the receptors and their downstream partners. Treatment of HER2 amplified subtype with monoclonal antibody trastuzumab has been effective. But the basal type is ER-, PR- and HER2- (triple negative breast cancer or TNBC). It is more frequent in young African

American women and the tumors are of high grade, mitotic, aggressive, show metastatic behavior and relapse with distant metastasis. TNBC is highly heterogeneous and in turn consists of six more subtypes and the lack of any targetable receptors leaves the systemic chemotherapy as the only treatment option.

Table 1.1. **Molecular features of the breast cancer subtypes.** Table taken from (TCGA network, 2012).

Table 1 | Highlights of genomic, clinical and proteomic features of subtypes

Subtype	Luminal A	Luminal B	Basal-like	HER2E
$ER^+/HER2^-$ (%)	87	82	10	20
$HER2^+$ (%)	7	15	2	68
TNBCs (%)	2	1	80	9
TP53 pathway	*TP53* mut (12%); gain of *MDM2* (14%)	*TP53* mut (32%); gain of *MDM2* (31%)	*TP53* mut (84%); gain of *MDM2* (14%)	*TP53* mut (75%); gain of *MDM2* (30%)
PIK3CA/PTEN pathway	*PIK3CA* mut (49%); *PTEN* mut/loss (13%); *INPP4B* loss (9%)	PIK3CA mut (32%) PTEN mut/loss (24%) INPP4B loss (16%)	PIK3CA mut (7%); PTEN mut/loss (35%); INPP4B loss (30%)	PIK3CA mut (42%); PTEN mut/loss (19%); INPP4B loss (30%)
RB1 pathway	Cyclin D1 amp (29%); *CDK4* gain (14%); low expression of *CDKN2C*; high expression of *RB1*	Cyclin D1 amp (58%); *CDK4* gain (25%)	*RB1* mut/loss (20%); cyclin E1 amp (9%); high expression of *CDKN2A*; low expression of *RB1*	Cyclin D1 amp (38%); *CDK4* gain (24%)
mRNA expression	High ER cluster; low proliferation	Lower ER cluster; high proliferation	Basal signature; high proliferation	HER2 amplicon signature; high proliferation
Copy number	Most diploid; many with quiet genomes; 1q, 8q, 8p11 gain; 8p, 16q loss; 11q13.3 amp (24%)	Most aneuploid; many with focal amp; 1q, 8q, 8p11 gain; 8p, 16q loss; 11q13.3 amp (51%); 8p11.23 amp (28%)	Most aneuploid; high genomic instability; 1q, 10p gain; 8p, 5q loss; *MYC* focal gain (40%)	Most aneuploid; high genomic instability; 1q, 8q gain; 8p loss; 17q12 focal *ERRB2* amp (71%)
DNA mutations	*PIK3CA* (49%); *TP53* (12%); *GATA3* (14%); *MAP3K1* (14%)	*TP53* (32%); *PIK3CA* (32%); *MAP3K1* (5%)	*TP53* (84%); *PIK3CA* (7%)	*TP53* (75%); *PIK3CA* (42%); *PIK3R1* (8%)
DNA methylation	–	Hypermethylated phenotype for subset	Hypomethylated	–
Protein expression	High oestrogen signalling; high MYB; RPPA reactive subtypes	Less oestrogen signalling; high FOXM1 and MYC; RPPA reactive subtypes	High expression of DNA repair proteins, PTEN and INPP4B loss signature (pAKT)	High protein and phospho-protein expression of EGFR and HER2

Percentages are based on 466 tumour overlap list. Amp, amplification; mut, mutation.

Whole exome sequencing of DNA from tumors of different breast cancer subtypes show recurrent somatic mutations in *PIK3CA*, *TP53*, *AKT1*, *GATA3* and *MAP3K1* with high rate of missense mutations particularly in *PIK3CA* and *TP53* (Banerji 2012). 30% of all breast cancer tumor subtypes carry mutations in *TP53*. The distribution of *TP53* mutations in the subtypes include 26% of luminal tumors (17% luminal A, 41% luminal B), 88% of basal-like carcinomas, and almost 50% of HER2+ tumors. There is high frequency of base substitution in luminal subtype while deletions and insertions are high in *TP53* in basal-like tumors (Bertheau 2013).

Clinical and prognostic value of TP53 somatic mutations in breast cancer

The mean age of tumor onset in patients with germline *TP53* mutations (Li-Fraumeni disease patients) is 24 years for carriers of missense mutations compared to 29 years for patients with nonsense mutation in *TP53*. With hotspot mutations R175H or R248W, mean age of tumor onset was between 15 and 20 years and 39 years with deletion in the *TP53* gene resulting in its loss-of-function. This indicates severe clinical outcomes associated with hotspot missense mutations in *TP53* with similar correlations for the somatic mutations (Hainaut 2016).

TP53 mutations and p53 protein overexpression are associated with prediction of response to chemotherapy. Failure of neoadjuvant treatment with fluorouracil, epirubicin and cyclophosphamide in breast cancer patients is attributed to both presence of *TP53* mutations and p53 overexpression captured by immunohistochemistry (IHC) (Kandioler-Eckersberger 2000). Mutations in p53 obtained from sequencing than IHC as the marker for mutation status has constantly been associated with poor prognosis in breast, colorectal, head and neck cancers and leukemia.

TP53 mutations are found to be more prevalent in recurring tumors compared to their corresponding primary breast tumors (41% vs 23%) indicating the importance of p53 mutations for tumor progression in breast cancer (Norberg 2001). Complete sequencing of the *TP53* coding regions from 316 breast cancer patients provided prognostic information about adjuvant systemic therapy and radiotherapy where the p53 mutations in the evolutionarily conserved regions (now known to harbor frequent or hotspot missense mutations) were associated with worse prognosis. Tumors that contained p53 mutations and were lymph-node positive did not benefit from adjuvant therapy with tamoxifen along with radiotherapy (Bergh 1995). In another study of 63 patients with locally advanced breast cancer and receiving doxorubicin monotherapy as the first line treatment showed primary resistance to doxorubicin therapy and early relapse in the breast cancer patients with mutation in *TP53* (Aas 1996). *TP53* mutations have also been linked to resistance to

anthracyclines and cisplatin-based chemotherapy. Investigation of the somatic mutations in *TP53* by exon sequencing in 1,794 primary breast cancer tumors with long-term follow up associated with aggressive phenotype like high tumor grade, large size, node positivity and low hormone receptors and high mortality rates. In fact, mutations in *TP53* and absence of progesterone receptor showed worst prognosis (Olivier 2006). *TP53* mutation status in 1,420 breast tumors (METABRIC cohort) by sequencing revealed different clinical relevance in subtypes of breast cancer. Mutations in p53 conferred worst overall breast cancer specific survival and marker of poor prognosis in ER+ patients. *TP53* mutation spectrum was different in different subtypes and linked to increased mortality in luminal B and HER2+ patients than luminal A and basal-like breast cancer patients (Silwal-Pandit 2014).

The advancement in whole genome or exome sequencing, DNA copy number changes, mRNA and protein expression has provided a comprehensive view of *TP53* mutations and their contexts where they appear. But one must be careful that prognostic and predictive significance of *TP53* mutations is quite variable and depends on the tumor type and the treatment and there is no universal clinical message deciphered from the *TP53* mutation status.

Other co-existing somatic mutations in breast cancer

The *TP53* mutations are early occurrences in breast cancer and a known clonal driver mutation that contributes to initiation and progression of the disease. But it exists with several other somatic mutations in the tumor resulting in genetic heterogeneity that complicates the treatment regime. The co-existing subclonal (less frequent, present in some cells not all) somatic mutations, the co-drivers, that drive tumorigenesis forward largely remain undiscovered. These functionally important driver mutations or as we introduced ‘the co-drivers’ can help to find out alternative direct drug targets (gene or proteins) or pathways for targeted therapies. For tackling the heterogeneous tumors, combination therapies would prove more beneficial than administration of a single drug.

Breast cancer is a heterogeneous disease with inter-tumor (between patients and within a patient) and intra-tumor (within a tumor from a patient) heterogeneity with respect to histological, prognostic and clinical aspects (Zardavas 2015). Intra-tumor heterogeneity has been observed at the spatial and temporal levels and is the consequence of tumor evolution resulting from selective pressure and this heterogeneity poses the biggest hurdle at developing targeted therapies for breast cancer. Analysis of different gene expression levels with microarray technology enabled the understanding of inter-tumor heterogeneity and classified breast cancer into subclasses- Luminal A, Luminal B, HER2 enriched and Basal like. Later, basal-like or TNBC was reported to be highly heterogeneous consisting of additional subtypes. With next-generation sequencing (NGS) of the breast tumors several genetic aberrations are found that affect the genomic landscape of breast cancer tumors namely, point mutations, small insertion or deletions (indels), amplification, chromosome or segmental duplication, translocations, inversions and fusions (Kalimutho 2019).

NGS identified high frequency of somatic mutations in *TP53, PIK3CA and GATA3* across all subtypes of breast cancer. The number of somatic mutations greatly vary between individual tumors. From whole-genome sequencing of 560 breast cancer tumors probable driver mutations, the highly mutated non-coding regions, genome rearrangements (tandem duplication or deletion) were identified that formed the somatic mutation landscape in breast cancer. *TP53, PIK3CA, MYC, CCND1, PTEN, ERBB2, ZNF703/FGFR1* locus*, GATA3, RB1* and *MAP3K1* are the most frequently mutated genes that account for 62% of drivers. Mutations in *TP53, MYC, PTEN* and *RB1* are more common in ER- breast tumors while mutations in *PIK3CA, GATA3, CCND1* and *MAP2K4* more common in ER+ breast tumors. Several other genes are mutated though at a much lower frequency and co-appeared with the above repertoire of frequently mutated genes (Nik-Zainal 2016). Mutations in genes *AKT2, ARID1B, CASP8*, *CDKN1B, MAP3K13, NCOR1, SMARCD1* and *TBX3* have been considered as other driver mutations in breast cancer (Stephens 2012) (Fig 1.13). For the pairwise association of somatic events tumors with inactivating mutations in *PTEN* also harbored recurrent *PIK3CA* mutations. Co-mutation of *TP53*

and *RB1* commonly occurs in TNBCs whereas *CDH1* and *PIK3CA* in lobular carcinomas (Pereira 2016).

Other than mutations in protein-coding regions, recurrent mutations are found in the non-coding RNA genes and promoters of genes like *FOXA1* (breast cancer oncogene) leading to its over-expression. Other recurrent somatic mutations in *AKT1* and *CBFB* and deletion in *RUNX1* and *MAGI3-AKT3* fusion enriched in TNBC were discovered from whole genome sequences of DNA from 103 breast tumors of different subtypes (Banerji 2012). From another study, deep sequencing of somatic mutations in 104 TNBC tumor samples identified varying clonal frequencies with more variation in basal subtype than non-basal TNBCs. Somatic mutations in *TP53, PTEN* and *PIK3CA* were clonally dominant while lower mutation frequencies were observed in cytoskeletal, cell shape and motility genes indicating their occurrence at a later stage during tumor progression (Shah 2012).

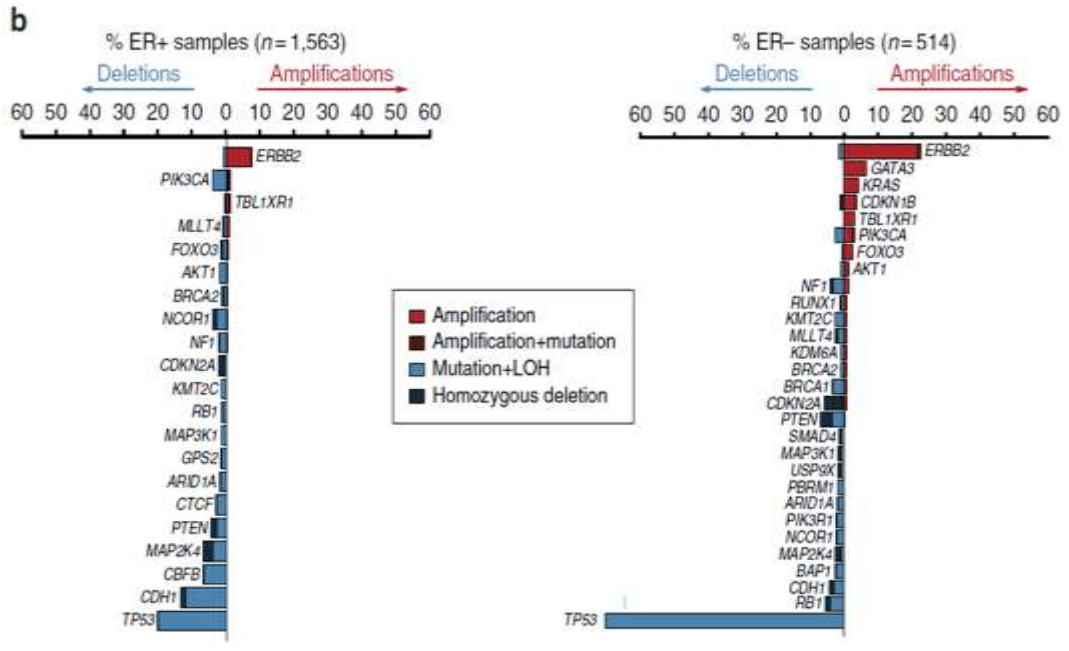


Figure 1.13. **Breast cancer tumors with copy number alterations in different genes in ER+ or ER- samples.** Image taken from (Pereira 2016).

Owing to breast tumor heterogeneity, combination therapies targeting multiple driver mutations may prove beneficial. The sequencing results from the breast tumor subtypes emphasize the importance of mutation-based stratification of breast cancer patients for designing therapeutic strategies.

The Breast Cancer Landscape

Breast cancer is a world-wide problem. It is the most common cancer found in women and it has been estimated that in the US in 2019, about 268,600 women will be diagnosed with breast cancer. In general, about one in every 8 women will be diagnosed with breast cancer in her lifetime. It is quite interesting to know that females on active military duties have up to 40% higher risk of breast cancer than the general population. Though breast cancer occurs 100 times more frequently in women, men can also be diagnosed with the disease. The race or ethnicity also has an influence on breast cancer incidence and mortality rates. In the US, there is enough variation and the Caucasians, and the Black Americans have higher incidence rates compared to Asians and Latinas. But mortality rates are highest among Black women. Triple Negative Breast Cancer (TNBC) is more frequent in African Americans typically the Basal-like subtype than the Caucasians. African Americans are also known to have shorter TTP (time to progression) and worse DFS (disease free survival) though no specific genomic profile, high prevalence genes or gene expression patterns in breast cancer biology are associated with racial differences (Ademuyiwa 2017).

Breast cancer is highly heterogeneous with major molecular subtypes luminal A, luminal B, HER2 overexpressing and basal-like. Luminal A subtype breast tumors are most common (70%) while the TNBC subtype tumors account for 15%. The TNBC and the HER2 overexpressing tumors have the worst prognosis. The metastatic disease is the cause for the estimated 90% deaths due to breast cancer and is not curable. Treatments like surgery, radiation, chemotherapy, hormone therapy, targeted therapy and immunotherapy are known to shrink the tumors and delay the

metastasis process. De novo and acquired resistance are the major challenges and no targeted therapies have been approved for the most aggressive TNBCs. The most commonly used combination chemotherapies show only 7-11% improvement in survival for younger women and just 2-3% for women above 50 years. The non-specific chemotherapy benefits only a small fraction of patients as we do not know how the heterogeneity of each tumor affects response to therapy or the recurrence.

Only very recently, the genome analysis methods are providing a comprehensive picture of genetic diversity in breast cancer tumors and could be used in routine clinical practice given that these methods are powerful, fast and reliable. Sequencing of TNBC tumors has revealed that copy number changes and mutation frequencies vary between and within tumors. Hence, the genome analysis method defines the landscape of this complex disease where there are few consistent and functionally characterized aberrations with hundreds of other genomic changes unique to individual tumor.

Cellular and molecular mechanism of action of mutant p53

The following sections provide a detailed read on how the mutant p53 protein gains oncogenic functions and how this phenomenon possibly shapes cancer development.

Hallmarks of cancer disrupted by mutant p53

Cancer progression in humans is a multistep process driven by genetic alterations that transform normal cells into malignant cells. These transformed cancer cells acquire certain molecular and biochemical capabilities that are shared by most cancer cells and these new traits dictate malignant growth. These cellular traits include- sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis and

activation of cell invasion and metastasis. These acquired capabilities comprise the ‘hallmarks of cancer’. Mutation in *TP53* directly affects cell viability and proliferation potential, apoptosis (programmed cell death) resistance, cell migration and invasion. There are two emerging hallmarks- reprogramming cellular metabolism and evasion of immune destruction. Genomic instability and inflammation supporting tumors have been described as the enabling hallmarks as genetic alterations drive tumor progression and sustained inflammatory responses often promote tumors. Thus, now there are a total of ten hallmarks of cancer that define the complexity of the neoplastic disease (Hanahan 2011).

Accumulation of mutant p53 protein in tumor cells

Accumulation of p53 protein in the nucleus is strongly associated with missense mutations in *TP53*. The tumor cells accumulate high but very heterogeneous level of mutant p53 proteins whereas the wildtype protein is barely detectable in cells under normal conditions. The protein levels build up only under stressed conditions. Concordance between *TP53* missense mutations in the DNA binding domain (DBD) determined by sequencing of cells from tumors and nuclear p53 mutant protein accumulation using immunohistochemistry in formalin fixed paraffin embedded tissue has been found in metastatic prostate cancer, invasive ductal carcinomas of breast, bladder cancer, colorectal adenocarcinomas, esophageal neoplasms, lung cancer and ovarian cancer that also strongly associates with poor prognosis. In fact, the ‘dominant negative’ effect or the ‘gain-of-function (GOF)’ effect of missense mutant p53 protein is attributable to its high concentration in the nucleus of the cancer cells and the formation of the hetero-tetramers.

In normal unstressed cells p53 protein accumulation is prevented by its efficient turnover and under stressed conditions modulation of these degradation pathways causes stabilization and accumulation of p53 protein leading to the activation of the target genes. Mdm2 is an essential component of the p53 degradation pathway which ubiquitinates p53 promoting degradation of p53 by proteasomes and Mdm2 also happens to be a p53 target gene. Mdm2 directly binds p53

at the N-terminal disrupting the binding of other transcriptional machinery components maintaining a very low p53 protein level under normal conditions. Mutant p53 proteins do not activate the Mdm2 expression, are unusually stable and accumulate to high levels in the tumor cells (Ashcroft 1999). Some cancer cell lines show hyperphosphorylation or hyperacetylation of mutant p53 protein that affects its stability and interaction with other proteins. This may change DNA binding and transcriptional regulation by mutant p53. As the p53 acetylation sites are shared for ubiquitylation, hyperacetylation impairs ubiquitylation and degradation of mutant p53 leading to its accumulation in the nucleus (Bode 2004). Overexpression of Mdm2 is common in many tumor types deactivating p53. Stabilization of mutant p53 is also supported by loss of INK4A expression and interaction with heat shock proteins HSP90 and HSP70 (Brosh 2009).

TP53 loss of function

The tumor suppressor gene *TP53* is addressed as the 'guardian of the genome' for the activation of senescence, cell-cycle arrest, apoptosis and DNA repair pathways for maintenance of genomic integrity under stressed conditions generated by DNA damage, oncogene activation, oxidative stress, hypoxia and nutrient deprivation. But after discovering *TP53*'s involvement in modulation of metabolic reprogramming, autophagy, tumor microenvironment signaling, inhibition of stem cell renewal and negatively regulating invasion and metastasis, *TP53* has now been designated the title of 'the master regulator'. For its broad range of tumor suppressive activities, without a doubt *TP53* remains the most mutated gene across all cancer types. *TP53* is inactivated either by missense or deletion mutations followed by loss of the remaining allele.

Homozygous deletion of p53 is seen in colon carcinomas, lung cancer, brain and breast tumors. The missense mutations concentrate in the highly conserved DNA binding domain of p53. The missense mutations in *TP53* occur in spontaneous tumors as well as the germline of Li-Fraumeni syndrome patients. The inherited mutations remain in the heterozygous state and pre-dispose an individual to carcinogenesis (Levine 1991). Though *TP53* undergoes deletion or truncation

mutations, missense mutations in the DBD of the protein are the most common mutations found in *TP53* in cancer patients. Typically, truncating or non-sense mutations, gene deletion and even missense mutations that abrogate binding of p53 to its consensus DNA sequences restrict the activation of the p53 target genes leading to its loss-of-function (LOF). Missense mutations in *TP53* give rise to full-length mutant p53 protein that interferes with the function of the wildtype p53. Oligomerization of the mutant and the wildtype p53 hinders sequence specific DNA binding and exerts the 'dominant negative effect' (DN). Loss of the other remaining wildtype *TP53* allele by mutation or deletion, enables 'gain-of-function' (GOF) properties to the mutant p53 with new functions that are independent of the wildtype p53 (Fig 1.14) (Brosh 2009).

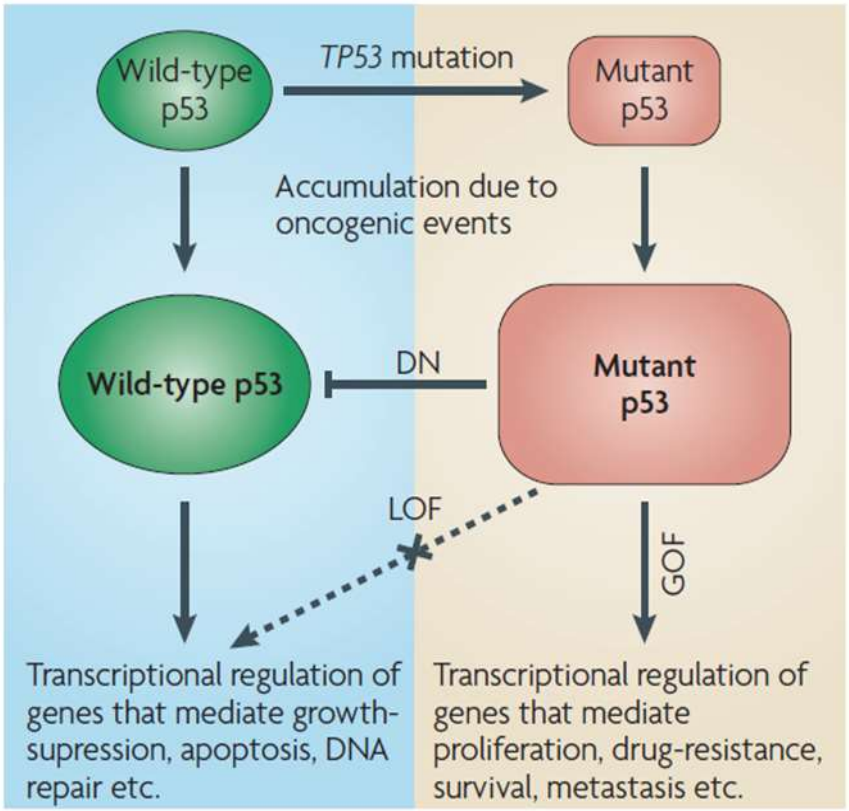


Figure 1.14. **Summary for phenotypic effects of *TP53* mutations.** Adapted from (Brosh 2009).

Mice homozygous for the null *TP53* allele develop normally but become susceptible to spontaneous tumors by 6 months since birth. These results show that the normal p53 is

dispensable for embryonic development, but its absence predisposes animals to neoplasms. (Donehower 1992). The p53 null homozygous mice particularly develop lymphomas and sarcomas. In comparison, heterozygous mice with one wildtype p53 allele rarely develop tumors before the age of 9 months. The homozygotes develop multiple types of cancer by 20 weeks including malignant lymphomas and sarcomas along with primary neoplasms of different cell-type of origin. The germline disruption of *TP53* show variable tumor latency and tissue distribution. $p53^{+/\Delta}$ serves as a model for Li-Fraumeni syndrome where the heterozygous mice develop sarcomas (57%), lymphomas (25%) and carcinomas. Determination of the p53 status in tumors of the heterozygous mice presented the complete absence of wildtype p53 allele by deletion rather than point mutation. The homozygous mice $p53^{\Delta/\Delta}$ have accelerated rate of tumorigenesis compared to heterozygous and p53 wildtype mice (Jacks 1994). Very few female mice with p53 null phenotype develop spontaneous mammary adenocarcinomas. Only tissue specific inactivation of p53 in a conditional mouse mammary tumor model developed spontaneous mammary tumors.

Loss of p53 accentuates tumor development whereas restoration of endogenous p53 expression result in regression of lymphomas and sarcomas (Ventura 2007) making *TP53* a perfect example of a tumor suppressor gene.

Dominant negative effect of mutant p53 protein

Acquisition of mutation in one *TP53* allele can interfere with the normal function of the remaining wildtype (WT) *TP53* allele. This phenomenon is referred to as the 'dominant negative' (DN) activity of the mutant *TP53*. Several possible reasons have been proposed to explain the ability of the dominant negative mutants to disrupt the WT allele function by 1) competition with the WT p53 where mutant p53 occupy the WT p53 binding sites and the co-factors 2) formation of hetero-tetramers comprising of both WT and mutant p53 monomers and 3) change in conformation of WT p53 by interaction with the mutant p53 into an inactive conformation suggestive of a prion-like

mechanism. Mutant p53 protein has no characteristics of an infectious protein and cannot be propagated, nullifying the prion like mechanism for dominant negative effect. The mechanism of this effect relies more on competition between mutant and wildtype p53. A higher concentration of mutant p53 can inhibit wildtype mediated transcription (Fig 1.15).

A yeast system shows that the p53 hotspot mutants exhibit dominant negative potential and this potential over WT p53 is based on formation of inactive hetero-tetramers and doesn't involve prion-like mechanism. Moreover, in this study the DN effect of p53 mutant R175H was dependent on its expression level and inhibited WT p53 in a dose-dependent manner. And inhibition of tetramerization abolished the dominant negative effect of the p53 mutants (Billant 2016). In an inducible system capable of expressing only mutant p53, only WT p53 or concurrent expression of both, p53 mutants R175H, R248W and R273H reduced the binding of WT p53 to its endogenous target genes p21, MDM2 and PIG3 reducing growth suppression, in presence of the mutant p53. The reduced ability of WT p53 to bind to the target genes in presence of mutant p53 was confirmed by chromatin immunoprecipitation (ChIP) demonstrating how the mutant p53 exhibit dominant negative effect (Willis 2004).

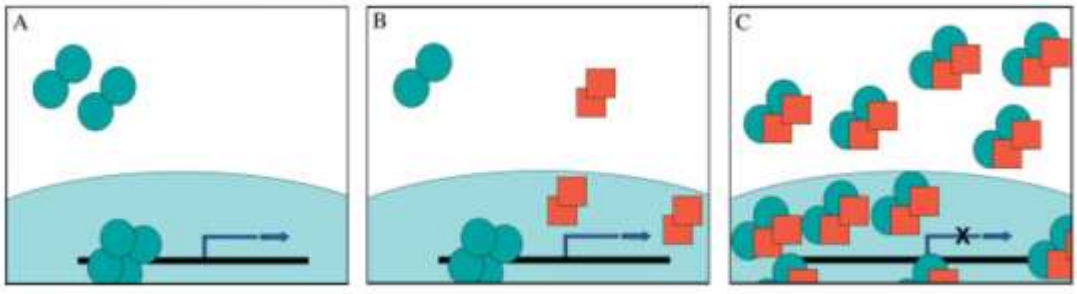


Figure 1.15. **Co-dominance of wildtype and mutant p53.** At higher concentration of mutant p53, hetero-tetramers of mutant and wild-type p53 are formed that poison the formation of wildtype p53 homo-tetramers. Image from (Goh 2011).

Several animal models expressing a mutant and a wildtype *TP53* allele demonstrated the dominant negative effect (DN) of mutant p53 over wildtype p53. The animals hemizygous for

endogenous wildtype p53 gene expressing multiple copies of mutant p53 Val135 had accelerated tumor growth and different tumor spectrum than the non-transgenic mice stressing on dominant negative effect of the mutant. These $p53^{+/-}$ transgenic mice developed lymphomas, soft tissue sarcomas, osteosarcomas and typically lung adenocarcinomas not seen in non-transgenic mice (Harvey 1995). In a conditional mouse model with heterozygous expression of mutant p53 in only the epithelium had increased incidence of spontaneous and UVB induced skin tumors. Reduced tumor latency time were observed in K14*Cre;p53*$^{R270H/+}$ (human R273H) mice compared to K14*Cre;p53*$^{+/+}$ mice. Upon UVB exposure, the tumor latency time of K14*Cre;p53*$^{R270H/+}$ mice were drastically reduced pointing towards the dominant negative effect of the R270H mutant equivalent to human R273H mutation (Wijnhoven 2007). Heterozygous mice $p53R^{172H/+}$ (human R175H) expressing mutant p53 to near wildtype levels had a different tumor spectrum from $p53^{+/-}$ mice. The $p53R^{172H/+}$ mice contracted more carcinomas and the osteosarcomas and the carcinomas metastasized where the osteosarcomas in $p53^{+/-}$ mice rarely metastasized (Liu 2000).

Mice with heterogeneous germline p53 loci $p53^{R270H/+}$ and $p53^{R172H/+}$ presented models for Li-Fraumeni syndrome and formed allele specific tumor spectra different from p53+/- mice showing that point/missense mutant p53 had oncogenic potential than simply loss of p53 function. $p53^{R270H/+}$ (human R273H) mice had increased tumor burden and incidence of carcinomas that were invasive. On the other hand, osteosarcomas were frequent in $p53^{R172H/+}$ (human R175H) with metastasis compared to $p53^{+/-}$. Mouse embryonic fibroblasts (MEFs) extracted from the $p53^{R270H/+}$ and $p53^{R172H/+}$ mice had a large fraction in the S-phase of cell cycle compared to $p53^{+/-}$ MEFs (Olive 2004). In a parallel study, osteosracomas and carcinomas in $p53^{+/515A}$ mice metastasized compared to the $p53^{+/-}$ heterozygous mice. Retrovirally expressed mutant *TRP53* (*TP53* in mice) in different genetic backgrounds- *Trp53-/-, Trp53*+/- and *Eμ-Myc/Trp53*+/+ demonstrated dominant negative effect of mutant *TRP53* in driving lymphomagenesis and a role of gain-of-function effect (GOF) of mutant *TRP53*. This system recapitulated the early stages of tumorigenesis in humans when a heterozygous state for the *TP53* exists before the loss of heterozygosity (LOH) at the *TP53* locus and mutant *TP53* driven cancers that retain a WT copy of

TP53. Among all the mutants R270H allowed the expression of WT *TRP53* target genes and displayed weakest dominant negative effects (Aubrey 2018). It is likely that the dominant negative effect exerted by the mutant p53 depends on a number of factors like mutant conformation, affinity of the mutant for WT p53 targets, DNA binding site, cell type and the available co-factors.

Gain of function (GOF) of neo-morphic mutant p53

The cancer associated *TP53* missense mutations not only abrogate the tumor suppressor functions of wildtype p53 but also endow the mutant proteins with new functions that actively contribute to the different stages of tumor progression. Hence, these 'neo-morphic' oncogenic properties acquired by the accumulation of abnormal amounts of mutant p53 protein in cell nuclei in absence of any wildtype p53 have been termed as the 'gain-of-function' (GOF). GOF has been studied by over-expressing a p53 mutant in a p53 null background in cells or in mice models to assess its impact on the cellular properties. Alternatively, by knocking-down endogenous mutant p53 by siRNA in tumor derived cells and monitoring the changes in cell phenotype has proven mutant p53 GOF (Oren 2010). Acquisition of missense mutations in *TP53* in human tumors is generally followed by loss of heterozygosity (LOH) at the other allele indicating a selective advantage of losing the wildtype allele after the other *TP53* allele is mutated. A strong selection for the mutant p53 proteins ascertains their role as tumor promoters instead of tumor suppressors. A pre-requisite for GOF is stabilization of mutant p53 and a recent work has shown that LOH is essential for mutant p53 stabilization and GOF.

Mouse tumors with high frequency of p53 LOH have accelerated tumor onset compared to $p53^{+/-}$ tumors and in the tumors with low frequency of LOH, $p53^{R248Q/+}$, mutant protein was not stabilized and hence, GOF was not observed (Alexandrova 2017). Over-expression of mutant p53 in cells with endogenous wildtype p53 do not qualify for GOF but dominant negative effect of the mutant. The biological manifestation of GOF of mutant p53 proteins in cultured cells encompass increased genomic instability, increased proliferation, apoptosis resistance, chemoresistance and

increased cell migration and invasion. Expression of mutant p53 proteins in murine fibroblast cell line (10)3 and human osteosarcoma cell-line SAOS-2 lacking endogenous wildtype p53 enhanced tumor formation in nude mice and plating efficiency in agar cell culture. The gain of tumorigenic potential of the mutants also activated multidrug resistance in the cells (Dittmer 1993). Missense mutant p53 proteins upregulate genes like *MDR1, PCNA, EGFR, c-MYC, cyclin A, cdk1, cdc25C* and repress the canonical WT p53 targets *p21, gadd45* and *PTEN*.

Expression of mutant p53 proteins in mouse models at the physiological levels and under physiological regulation cleared any doubt about the cancer associated features of mutant p53 GOF. Mutant p53 knock-in mice with hotspot mutations R248Q and G245S showed early onset of different tumor types in comparison to p53 null mice. Mice with R248Q/- had expanded hematopoietic and mesenchymal stem cells and R248Q/R248Q mice had broadened tumor spectrum with enhanced Akt signaling compared to null mice (Hanel 2013). $p53^{R270H/-}$ and $p53^{R172H/-}$ (equivalent to human R273H and R175H) mice developed tumor spectrum specific to

Table 1.2. **Some of the mutant p53 protein interacting partners.** Table from (Freed-Pastor 2012).

Protein	Interaction with wild-type p53	p53 mutant(s)	Outcome(s)	References
NF-Y	Yes	R175H, R273C	Transactivation of NF-Y target genes (cell cycle progression)	Di Agostino et al. 2006 Liu et al. 2011
Sp1	Yes	V134A, R175H, R249S, R273H	Transactivates Sp1 target genes	Bargonetti et al. 1997; Chicas et al. 2000; Torgeman et al. 2001 Hwang et al. 2011
Ets-1	Yes	V143A, D281G	Transactivates Ets-1 target genes (i.e., MDR-1)	Sampath et al. 2001; Kim et al. 2003
VDR	Yes	R175H	Transactivates VDR target genes to promote cell survival	Stambolsky et al. 2010
SREBP-2	Unknown	R273H, R280K	Transactivates SREBP target genes	Freed-Pastor et al. 2012
TopBP1	Yes	V143A, R175H, R248W, R248Q, R249S, R273H, R273C	Mediates p53 interaction with NF-Y	Liu et al. 2011
Pin1	Yes	R280K	Promotes mutant p53 gain of function	Girardini et al. 2011
MRE11	No	R248W, R273H	Promotes genomic instability	Song et al. 2007
PML	Yes	R175H, R273H	Enhances mutant p53 transcriptional activity	Haupt et al. 2009
p63	No	R175H, Y220C, R248W, R273H (not D281G)	Inhibits p63-mediated transcription of p53 target genes	Gaiddon et al. 2001; Strano et al. 2002
p73	No	R175H, Y220C, V143A, R248W (not R273H)	Inhibits p73-mediated transcription of p53 target genes	Di Como et al. 1999; Marin et al. 2000; Gaiddon et al. 2001

the allele compared to p53 null mice including frequent metastatic carcinomas and endothelial tumors (Olive 2004).

Several models for mechanism of GOF for mutant p53 proteins have been proposed. For GOF the different missense mutant p53 proteins interact with many other proteins to either enhance or inhibit the activities (Table 1.2). Some of the proposed mechanisms include 1) Specific p53 mutant binding elements or regions in DNA but as mutant p53 loses sequence specific DNA binding it is unlikely that sequence specific mutant p53 response elements exist. Rather, mutant p53 could sense DNA in a structure specific manner (Freed-Pastor 2012). 2) Mutant p53 interacts with other transcription factors in a complex with co-factors and other proteins resulting in activation of their target genes. Mutant p53 interacts with transcription factor NF-Y upregulating expression of cyclin A, cyclin B1, cdk1 and cdc25C post DNA damage. Following DNA damage mutant p53 binds to NF-Y target promoters and recruits p300 for histone acetylation (Di Agostino 2006). Mutant p53 enhances the expression of NF-Y, Sp1, Ets1 and VDR target genes all known to be repressed by the wildtype version of p53. For example, p53 mutants commonly found in breast cancer cell lines bind and upregulate chromatin methyltransferases MLL1 and MLL2 and acetyltransferase MOZ increasing genome wide histone methylation and acetylation culminating into uncontrolled cell proliferation. MLL1, MLL2 and MOZ are specifically upregulated in patient derived p53 mutated breast tumors and not in p53 wildtype or null tumors. Genetic knock-down or inhibition of MLL1 methyltransferase complex markedly lowered cancer cell proliferation (Zhu 2015). 3) Mutant p53 can form aggregates with other proteins decreasing binding of transcription factors and co-factors to DNA. Murine p53R172H (human R175H) mutant binds to p63 and p73 and inhibits their function as transcription factors (Lang 2004). 4) Mutant p53 interacts with proteins other than transcriptional regulators and either enhances or blocks the protein function (Muller 2013) (Fig 1.16). Generation of humanized p53 mutant knock-in (HUPKI) mouse models with common p53 mutants R273H, R248W and R175H showed interaction of mutant p53 with Mre11 (part of MRN complex) inactivating ATM dependent DNA damage response resulting in chromosomal translocation and disrupted G_2/M checkpoint. This formed the basis for common

gain of function of the mutants by inducing genetic instability (Song 2007, Liu 2010). 5) Mutant p53 may introduce alterations in chromatin architecture by making portions of the genome accessible or not accessible to transcription factors creating broad changes in gene expression. The GOF effect of mutant p53 proteins are likely to be context specific and dependent on type of cancer, cell type and cellular microenvironment. The difference in the mutant p53 protein conformations (DNA contact vs structural described in an earlier section) could also determine the protein-protein interactions and binding affinities (Kim 2018).

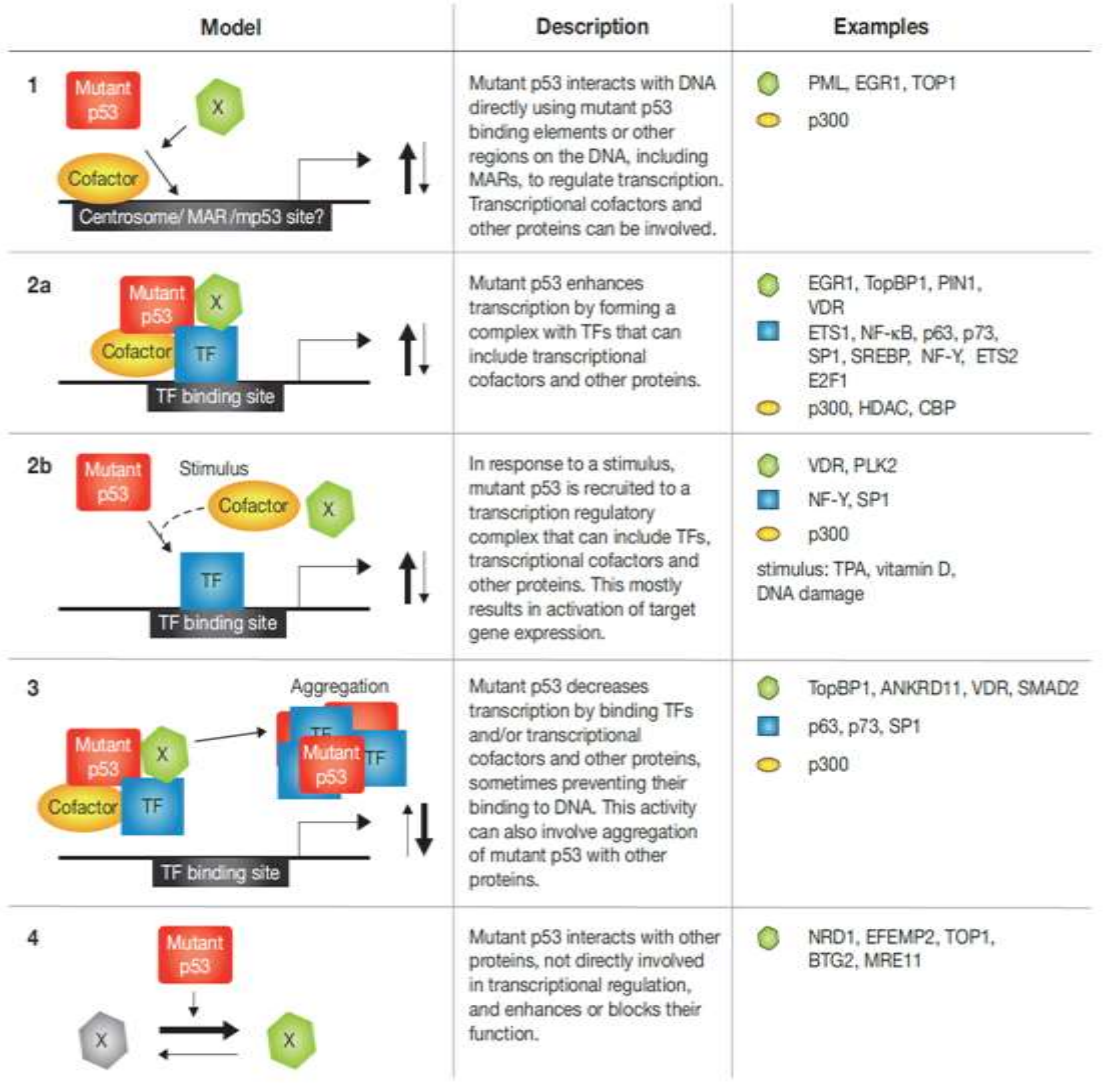


Figure 1.16. **Proposed mechanisms for explaining GOF of mutant p53 proteins.** Image adapted from (Muller 2013).

Cancer is a multistep process

Cancer tumors evolve from a benign state to the malignant form by accumulation of a large number of mutations. Some of these mutations provide a selective advantage to the growth of tumor while most of them remain neutral. The advantageous mutations could be recurrent mutations or infrequent mutations. The molecular interactions between the frequent (clonal) and the infrequent (subclonal) mutations shape the molecular heterogeneity in the tumors. A difficult task is to recognize these less prevalent mutations from the neutral mutations and an urgent need.

Clonal evolution in cancer

In the 1950s Armitage and Doll proposed the multistage model of carcinogenesis and in 1976, Peter Nowell proposed the clonal evolution theory for the tumor development. Nowell's hypothesis of tumor evolution described that tumor formation is initiated by a neoplastic change in a normal cell that acquires a selective growth advantage over the other normal cells. As time lapses genetic instability increases and mutations accumulate, and more cell subpopulations arise. Sequential selection by an evolutionary process forms a malignant tumor with high genetic diversity rich in genetic structural variations and point mutations (Nowell 1976). These proposals formed the basis for understanding the process of tumor initiation and progression to malignancy. The proposals have been supported by the study of progression of colorectal cancers that undergo histopathological changes from early adenomas to carcinomas with sequential acquisition of mutations and clonal expansions and breast cancers that proceed from ductal carcinoma in-situ (DCIS) to invasive carcinomas with increasing mutation burden and genetic alterations (Vogelstein 1993, Fujji 1996). In simple words a growing cancerous tumor can be viewed as entities evolving as per Darwinian rules. The tumors accumulate mutations and the selective pressure from the tumor environment select beneficial mutations that drive the expansion of subclones, and different subclones together form the tumor mass.

The clonal evolution is a dynamic process where the tumor cells constantly acquire new somatic mutations or aberrations in a changing microenvironment giving rise to a heterogeneous population of cells and a positive selection of phenotypic traits increases the adaptability of the tumors and their aggressiveness (Sun 2018). The chromosomal instability of the tumor cells is crucial for cancer evolution as it alters the mutation rate resulting in accumulation of adaptive but grossly genetically altered clones increasing the tumor cell population diversity.

Genomics has proved to be a powerful method to quantitatively measure evolving clones in tumors in space and time. Next generation sequencing (NGS) techniques have provided a detailed map of the genetic landscape of different cancer types and highlight the tumor evolutionary patterns. These techniques no longer support the linear evolution model which supports that a single clone survives over time and mutations are acquired sequentially. The NGS data rather supports the branching evolution of the tumors as continuous mutation accumulation produce genetic divergence. The high genetic diversity of the tumors has prompted the existence of neutral evolution within tumor cell population that occurs in absence of selection. Thus, subclonal composition is preserved over time contributing more to tumor heterogeneity (Turajlic 2019). The high genomic instability often results in loss, gain, translocation, rearrangement or fusion of chromosomes causing rapid adaptive evolution than gradualism called punctuated evolution.

The continuous acquisition of genetic mutations in a tumor makes intratumor genetic heterogeneity unavoidable in cancer tumors and amplifies genomic differences among single cells in a tumor. Thus, each cell remains as a record for the distinct mutational events shaping the evolutionary history. Bulk sequencing of cells captures the highly prevalent or recurrent mutations but miss the low frequency variants. It also doesn’t provide any information on co-occurrence of the mutations. On the other hand, single cell sequencing provides very high resolution and traces events in each cell. Single cell sequencing resolves tumor genetic diversity, deep sequencing reveals population dynamics and mutation frequencies and multi-region sequencing from a tumor infers tumor evolution patterns (Williams 2019).

Taking samples from a tumor separated in space or time is the most direct and simplest approach to understand tumor evolution. Comparing the genome at different stages of tumor development exhibits the temporal sequence of events identifying the early and the late events with the clear picture of the mutational landscape. Whole genome and exome single cell sequencing approach of single normal and tumor nuclei from an ER+ and TNBC revealed that point mutations evolved gradually establishing large clonal diversity. In contrast, single nuclei copy number profiling showed that chromosomal rearrangements occurred early during tumor progression and remained stable throughout tumor evolution (punctuated evolution). This method of single cell sequencing could capture mutations occurring at low frequencies missed in bulk sequencing of tumor masses. They also deduced that TNBC tumors have increased mutation rate, 13 times more than ER+ tumors, explaining the high heterogeneity observed in the TNBC subtype compared to the other breast cancer subtypes (Wang 2014).

Sequencing a set of genes or whole genome sequencing of locally relapsed or metastatic breast cancer and the primary tumors indicated that metastatic and relapsed clones disseminated late from the primary tumors and acquired new mutations not seen in the primary tumor. These metastatic drivers or drivers in the local relapses (mostly point mutants) were separate from the early drivers. Mutation in the less frequent cancer genes *SWI/SNF, ARID1A, ARID1B* were found only in the recurrent tumor. Sequencing of the local recurrences and distant metastasis included very different pathways important for designing therapies (Yates 2017).

Breast cancer is not a single disease entity but has distinct subtypes with different clinical outcomes. The molecular subtypes luminal A, luminal B, HER2+ and basal-like each have subtype specific genetic alteration patterns and possibly explain the observed tumor heterogeneity (described in section ‘other co-existing somatic mutations in breast cancer’). The breast tumor initiation and progression are driven by genetic alterations and evolves into the malignant form. Identification of subtype specific mutation patterns and altered molecular pathways is required for the successful development of targeted therapy regimes. Application of

single-cell DNA and RNA sequencing accompanied by bulk exome sequencing traced the evolutionary trajectory of clones in TNBC tumors. TNBC tumors extracted from patients undergoing neoadjuvant chemotherapy showed that resistant clones pre-existed and were selected by chemotherapy and the transcriptional profiles of resistant signatures resulted from reprogramming in response to chemotherapy (Kim 2018).

The somatic driver and passenger mutations in cancer

Development of a cancer tumor is an example of an evolutionary process where the growing tumors accumulate mutations that promote genetic diversification and selection of clones in the tumor microenvironment result in clonal expansion and the final architecture of the tumor (Greaves 2012). The tumor evolution occurs through the interaction between the functionally advantageous somatic 'driver' mutations and the neutral 'passenger' mutations. The driver mutations are identified as recurrent mutations seen in multiple cancers, present in greater proportion of the tumor, mutation rate much higher than the normal cell background mutation rate, associated with clonal expansions and involved in pathways leading to oncogenesis particularly providing a growth advantage to the tumor. The drivers may bear missense, non-sense, frameshift or deletion mutations. The driver mutations could be in the gene coding regions, in non-coding regions, miRNAs and in the epigenome. It has been suggested that driver mutations provide an average fitness advantage of only 0.4% indicating that the drivers help in the tumor initiation stages and requires interaction with other less prevalent somatic mutations that we introduced as the co-drivers for tumor maintenance and spread.

A central topic of debate is how many driver mutations are necessary for the onset of tumorigenesis. In a recent study, application of mathematical methods for cancer mutation network analysis from 7665 tumors and 30 cancer types identified that 27% of the tumors could be placed into modules linked to specific driver mutations and the rest belonged to a diffused network with lower specificity of driver mutations. The mean number of driver mutations needed

for cancer onset was suggested with a range from 1 to 5. The cancers placed in the diffused component had late onset and suggested presence of multiple unidentified or interchangeable driver mutations (Iranzo 2018). The number of driver mutations increase with increase in the mutation burden, but this increase is not linear. The driver genes comprise of 1-4% of all genes, 97-98% passenger genes and less than 0.5% genes are lost in cancer (Martincorena 2017). The passenger mutations do not provide any positive or negative selective growth advantage to the tumor cells and remain in the tumor mass by chance during cell division and clonal expansion.

The classical model of cancer progression describe that a handful of recurrently mutated driver genes promote tumor development while the passenger mutations do not participate in the process. Initiation of cancer therapy often gives rise to resistant tumors that relapse. It is possible that the existing passenger mutations under the selective pressure of therapy get converted into minor drivers, clonally expand and cause recurrence of tumors. These resistance mutations may not provide a strong proliferative advantage but offer advantage of tumor persistence (Stratton 2009).

In most cancers, few frequent driver mutations co-exist with several low-frequency mutations that may have biological significance and weak tumor promoting effects by themselves. The number of passenger mutations far exceed the number of driver mutations. Such less frequently mutated genes have also been termed as the ‘mini-drivers’ and several of these mutations may equal for a major strong driver change in the context of genome instability as observed in the cancerous tumors (Castro-Giner 2015). They provide small selective advantage to the tumor but may not be essential for cancer growth. These rare ‘candidate drivers’ genes may harbor missense or small indel mutations, DNA copy number variations and even gene fusions and be located at non-coding genomic regions. The infrequently mutated genes cannot be identified using the usual method of frequency determination as for the strong frequently mutated driver genes (Pon 2015). Computational methods to identify such low-frequency drivers from sequencing data of tumors have considered mutation frequency, mutation clustering pattern and mutations at conserved

sites that have not been considered before. But an unbiased method to identify the rare genetic drivers is functional validation. Methods like CRISPR-Cas9 that can introduce mutations in untransformed cells, RNAi and drug screening on established cancer cells are valuable tools to assay functional consequences of different mutations (Scholl 2019).

The breast cancer genomic landscape consists of few frequently mutated genes, the 'mountains' and a large number of genes mutated at a low frequency called the 'hills' (Wood 2007). Both the hills and the mountains provide fitness to the growing breast tumors (Fig 1.17). The tumor suppressor gene *TP53* is the most frequently mutated gene in breast cancer and is an established driver gene. Accumulation of missense mutant p53 is recognized in early phase of breast cancer (in-situ ductal carcinomas) and is maintained through invasive carcinomas and lymph node metastases (Davidoff 1991). The *PIK3CA* gene is the next recurrently mutated gene. A key challenge is to distinguish the drivers from the passengers. We have applied an efficient function-based approach to separate the rare drivers from the passenger mutations using an

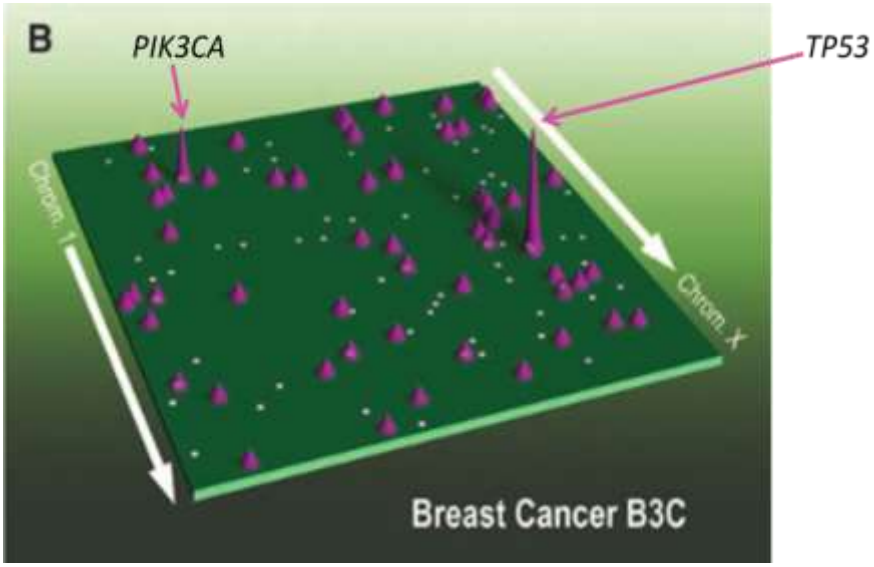


Figure 1.17. **The breast cancer genome landscape.** *TP53* and *PIK3CA* are recurrently mutated in breast cancers and are the recognized driver genes. The tiny purple hills represent the less frequently mutated candidate driver genes and the white dots are the neutral passenger mutations. Image from (Wood 2007).

efficient model to introduce genetic alterations in non-transformed cells and the systematic analysis of the phenotype change in the cells.

Cancer tumor heterogeneity

Tumor heterogeneity is driven by genetic alteration due to genome instability and selection of advantageous genotype during tumor evolution and poses a threat for targeted therapies. Tumors originating from different tissues and cell types or tumors from same tissue and cell type differing in genomic aberration, phenotype and drug sensitivity comprise the inter-tumor heterogeneity. Within a tumor, distinct cell subpopulations arise by genetic variations and selection of cells with a phenotype advantage in the given context resulting in intra-tumor heterogeneity (Fig 1.18). These subpopulations may intermingle or may remain separated as clones in different regions of the tumor (Burrell 2013).

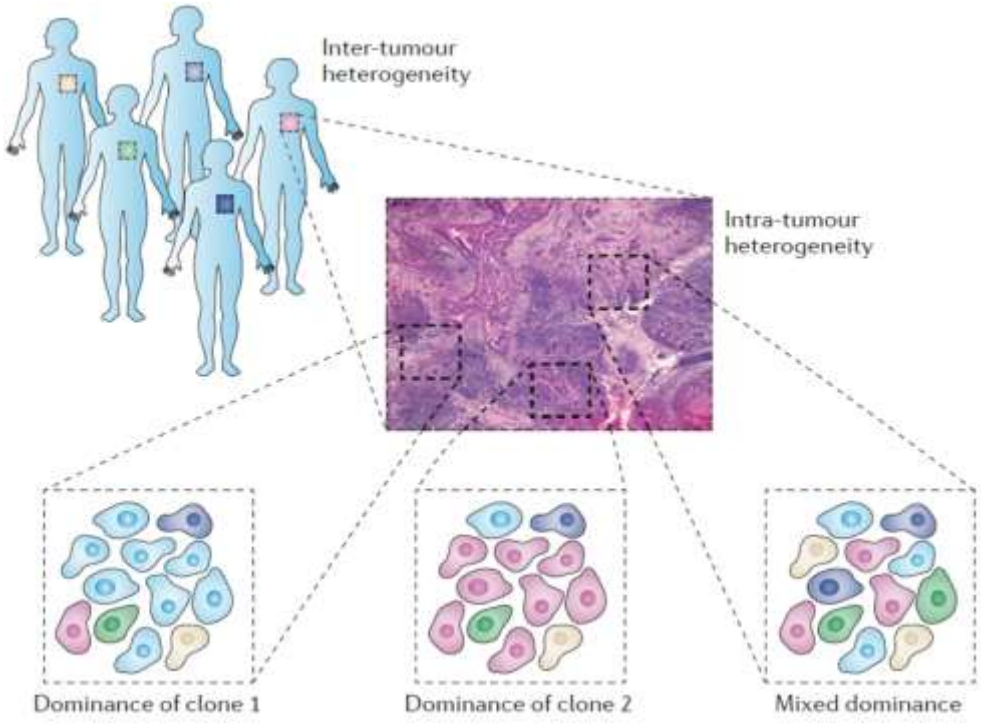


Figure 1.18. **Inter and intra-tumor heterogeneity.** Tumor heterogeneity result from genetic and non-genetic influences and complicate therapeutic decisions. Image taken from (Marusyk 2012).

Before the widespread use of next generation sequencing (NGS) methods, first proof of presence of multiple subpopulation of cells with differing genotypes in a tumor came from karyotyping and fluorescent in-situ hybridization (FISH) that highlighted patterns of chromosomal rearrangement and copy number aberrations (Fisher 2013). Deep exome sequencing, whole genome, multi-region and single cell sequencing have proved to be powerful tools to detect intra-tumor heterogeneity and added overflowing information about point mutations and small indels. Application of whole genome and targeted sequencing of breast cancer samples from multiple regions in tumors elucidated the subclonal diversification of the tumors. The commonly mutated breast cancer genes (clonal or frequent mutations) *TP53, PIK3CA, PTEN, BRCA* and *MYC* appeared along with subclonal mutations (mutations present in small fraction of cells in the tumor, co-drivers) with implication in resistance to chemotherapy (amplification of *CDK6, FGFR2*, deletion in *RUNX1*) and acquisition of invasive potential (Yates 2015).

The phenotypic heterogeneity of a tumor is determined by increased genomic instability of the cancer cells that leads to higher mutation rates and acquisition of diverse mutations and genotypes. Apart from genetic changes epigenetic aberrations (DNA methylations, chromatin remodeling, post-translational modification of histones) and heterogeneous tumor microenvironment also affect the clonal composition of the tumors. All these factors give rise to tumors with distinct clonal architecture which is spatially and temporally heterogeneous (Marusyk 2012). The heterogeneous tumors act like complex systems where the minor subpopulations affect tumor growth and maintain tumor heterogeneity along with the dominant clones. It is not easy to filter out passengers from the co-drivers but genome sequencing with functional screens in physiologically relevant background may identify clinically important minor subclonal mutations.

Tumor heterogeneity is a major source of therapeutic resistance and is a marker for poor clinical outcome. The tumor evolution following the branching evolution and the clonal composition of the tumors changes over time during tumor progression adding to the layer of complexities. Targeted therapies tend to exploit the dependence of tumors on clonally dominant driver mutations. But

such strategies help only a fraction of patients and those who benefit, the outcome lasts for a limited duration and is often accompanied by relapse and such clinical outcomes could be explained by the presence of minor subclones or co-drivers resulting in tumor heterogeneity. Therapeutic decisions based on determining the dominant clone in a tumor would be ineffective if clinically and biologically important co-drivers are not accounted for as such subclones can select for drug resistance and relapse. The intra-tumor heterogeneity provides plasticity and diversity to the tumor to adapt under varying microenvironmental pressure. Hence, there is a need to explore the relevant minor or subclonal drivers, the co-drivers, and their implication in therapeutic outcome. The heterogeneous or polyclonal tumors can be tackled by targeting multiple pathways simultaneously or sequentially. Targeting trunk mutations first and then blocking the emerging subclones could prove to be an ideal treatment combination. However, combinatorial targeted therapeutic strategies are associated with high toxicity and costs and often limit the use of such strategies (McGranahan 2015).

Breast cancer evolves from a transformed normal cell via hyperplasia, premalignant tumor, in-situ carcinoma to invasive carcinoma by accumulation of various genetic alterations (Beckmann 1997). Consequentially breast cancer displays heterogeneity with inter and intra-tumor diversity that influences breast cancer progression and therapy resistance (Polyak 2011). Gene expression analysis of breast tumors have identified the major subtypes- luminal, HER2 overexpressing and basal-like. Each subtype has different genetic constitution, risk factor, response to treatment, disease progression and pattern of metastasis. Luminal tumors are ER+ and PR+ and are treated with hormonal therapies. HER2 amplified tumors respond well to anti-HER2 antibody trastuzumab. But there is no targeted therapy for basal-like subtype due to lack of the targetable receptors ER, PR and HER2 (triple negative). Systemic chemotherapy is the only available option and unfortunately, only 20% of tumors respond to the standard chemotherapy. Triple negative breast cancer (TNBC) is more complex than was thought before where gene expression analysis identified six more of its subtypes, each with unique molecular markers. The

six subtypes include basal-like (BL1 and BL2), immunomodulatory, mesenchymal (M), mesenchymal stem-like (MSL) and luminal androgen receptor (LAR) subtype (Mayer 2014).

Owing to this heterogeneity, no single targeted therapy has been approved for treatment of TNBC. Single targeted therapies including platinum agents, PARP inhibitors, PI3K and MEK inhibitors did not bear favorable outcomes but combination of two or more agents would make a more rational approach to treat TNBC. The TNBC tumors contain more genetic mutations compared to luminal and HER2+ subtype due to high mutation rate and identification of the co-drivers and designing targeted therapies against them could be effective in treating the TNBC patients.

Hypothesis and approach

Unlike the established tumor suppressor genes *RB1* and *APC*, 80% of mutations in *TP53* are missense mutations instead of indels, non-sense and frameshift mutations. These missense mutations in p53 are associated with heterogeneous phenotypes because of the gain-of-function (GOF) or neo-morphic activities with the recognition of missense mutant p53 as oncogenic proteins. For such an atypical tumor suppressor, some open questions remain that need to be answered.

1. How are these p53 mutant protein variants functionally important in the context of breast cancer? Are the proteins functionally equal?
2. How are these diverse mutant variants contributing to the heterogeneity of malignant breast tumors like TNBC? Do they have any other accomplices?

The dissertation provides an answer to these questions that would help in achieving a better understanding of how the missense mutant p53 proteins drive breast tumor progression, tumor heterogeneity, how breast cancers become metastatic and may build a scope for developing targeted therapies to counteract the tumor progression.

We hypothesized that the TNBC tumor heterogeneity is attributed to 1) phenotypic or functional difference across the different *TP53* missense mutations themselves as well as 2) and the functional crosstalk with the distinct co-driver mutations and pathways unique to each *TP53* mutant.

We addressed these questions by 1) detailed phenotypic and molecular profiling of a panel of clinically important *TP53* mutations found in breast cancer and 2) adopted a functional genomics approach to identify the co-driver candidates that promoted cell invasion. In the following chapters we describe all our approaches and the results obtained in detail.

References

Aas, Turid, Anne-Lise Børresen, Stephanie Geisler, Birgitte Smith-Sørensen, Hilde Johnsen, Jan E. Varhaug, Lars A. Akslen, and Per E. Lønning. "Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients." *Nature medicine* 2, no. 7 (1996): 811.

Ademuyiwa, Foluso O., Yu Tao, Jingqin Luo, Katherine Weilbaecher, and Cynthia X. Ma. "Differences in the mutational landscape of triple-negative breast cancer in African Americans and Caucasians." *Breast cancer research and treatment* 161, no. 3 (2017): 491-499.

Bailey, Matthew H., Collin Tokheim, Eduard Porta-Pardo, Sohini Sengupta, Denis Bertrand, Amila Weerasinghe, Antonio Colaprico et al. "Comprehensive characterization of cancer driver genes and mutations." *Cell* 173, no. 2 (2018): 371-385.

Baker, Suzanne J., Eric R. Fearon, Janice M. Nigro, A. C. Preisinger, J. M. Jessup, D. H. Ledbetter, D. F. Barker, Y. Nakamura, R. White, and B. Vogelstein. "Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas." *Science* 244, no. 4901 (1989): 217-221.

Banerji, Shantanu, Kristian Cibulskis, Claudia Rangel-Escareno, Kristin K. Brown, Scott L. Carter, Abbie M. Frederick, Michael S. Lawrence et al. "Sequence analysis of mutations and translocations across breast cancer subtypes." *Nature* 486, no. 7403 (2012): 405.

Beckerman, Rachel, and Carol Prives. "Transcriptional regulation by p53." *Cold Spring Harbor perspectives in biology*2, no. 8 (2010): a000935.

Beckmann, M. W., Dieter Niederacher, Hans-Georg Schnürch, Barry A. Gusterson, and Hans Georg Bender. "Multistep carcinogenesis of breast cancer and tumour heterogeneity." *Journal of molecular medicine* 75, no. 6 (1997): 429-439.

Bergh, Jonas, Torbjörn Norberg, Sigrid Sjögren, Anders Lindgren, and Lars Holmberg. "Complete sequencing of the p53 gene provides prognostic information in breast cancer patients, particularly in relation to adjuvant systemic therapy and radiotherapy." *Nature medicine* 1, no. 10 (1995): 1029.

Bertheau, Philippe, Jacqueline Lehmann-Che, Mariana Varna, Anne Dumay, Brigitte Poirot, Raphaël Porcher, Elisabeth Turpin et al. "p53 in breast cancer subtypes and new insights into response to chemotherapy." *The Breast* 22 (2013): S27-S29.

Bianchini, Giampaolo, Justin M. Balko, Ingrid A. Mayer, Melinda E. Sanders, and Luca Gianni. "Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease." *Nature reviews Clinical oncology* 13, no. 11 (2016): 674.

Bieging, Kathryn T., Stephano Spano Mello, and Laura D. Attardi. "Unravelling mechanisms of p53-mediated tumour suppression." *Nature Reviews Cancer* 14, no. 5 (2014): 359.

Birch, Jillian M., Valerie Blair, Anna M. Kelsey, D. Gareth Evans, Martin Harris, Karen J. Tricker, and Jennifer M. Varley. "Cancer phenotype correlates with constitutional TP53 genotype in families with the Li–Fraumeni syndrome." *Oncogene* 17, no. 9 (1998): 1061.

Bode, Ann M., and Zigang Dong. "Post-translational modification of p53 in tumorigenesis." *Nature Reviews Cancer*4, no. 10 (2004): 793.

Børresen-Dale, Anne-Lise. "TP53 and breast cancer." *Human mutation* 21, no. 3 (2003): 292-300.

Burrell, Rebecca A., Nicholas McGranahan, Jiri Bartek, and Charles Swanton. "The causes and consequences of genetic heterogeneity in cancer evolution." *Nature* 501, no. 7467 (2013): 338.

Cancer Genome Atlas Network. "Comprehensive molecular portraits of human breast tumours." *Nature* 490, no. 7418 (2012): 61.

Carey, Lisa, Eric Winer, Giuseppe Viale, David Cameron, and Luca Gianni. "Triple-negative breast cancer: disease entity or title of convenience?." *Nature reviews Clinical oncology* 7, no. 12 (2010): 683.

Castro-Giner, Francesc, Peter Ratcliffe, and Ian Tomlinson. "The mini-driver model of polygenic cancer evolution." *Nature Reviews Cancer* 15, no. 11 (2015): 680.

Chen, Sidi, Neville E. Sanjana, Kaijie Zheng, Ophir Shalem, Kyungheon Lee, Xi Shi, David A. Scott et al. "Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis." *Cell* 160, no. 6 (2015): 1246-1260.

Cho, Yunje, Svetlana Gorina, Philip D. Jeffrey, and Nikola P. Pavletich. "Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations." *Science* 265, no. 5170 (1994): 346-355.

Chow, Ryan D., and Sidi Chen. "Cancer CRISPR screens in vivo." *Trends in cancer* 4, no. 5 (2018): 349-358.

Davidoff, Andrew M., Billie-Jo M. Kerns, J. Dirk Iglehart, and Jeffrey R. Marks. "Maintenance of p53 alterations throughout breast cancer progression." *Cancer Research* 51, no. 10 (1991): 2605-2610.

Donehower, Lawrence A., Michele Harvey, Betty L. Slagle, Mark J. McArthur, Charles A. Montgomery Jr, Janet S. Butel, and Allan Bradley. "Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours." *Nature* 356, no. 6366 (1992): 215.

Eliyahu, Daniel, Avraham Raz, Peter Gruss, David Givol, and Moshe Oren. "Participation of p53 cellular tumour antigen in transformation of normal embryonic cells." *Nature* 312, no. 5995 (1984): 646.

Finlay, C. A., P. W. Hinds, T. H. Tan, D. Eliyahu, M. Oren, and A. J. Levine. "Activating mutations for transformation by p53 produce a gene product that forms an hsc70-p53 complex with an altered half-life." *Molecular and cellular biology* 8, no. 2 (1988): 531-539.

Finlay, Cathy A., Philip W. Hinds, and Arnold J. Levine. "The p53 proto-oncogene can act as a suppressor of transformation." *Cell* 57, no. 7 (1989): 1083-1093.

Fischer, M. "Census and evaluation of p53 target genes." *Oncogene* 36, no. 28 (2017): 3943.
Fisher, R., L. Pusztai, and C. Swanton. "Cancer heterogeneity: implications for targeted therapeutics." *British journal of cancer*108, no. 3 (2013): 479.

Fujii, Hiroaki, Carla Marsh, Paul Cairns, David Sidransky, and Edward Gabrielson. "Genetic divergence in the clonal evolution of breast cancer." *Cancer Research* 56, no. 7 (1996): 1493-1497.

Gao, Ruli, Alexander Davis, Thomas O. McDonald, Emi Sei, Xiuqing Shi, Yong Wang, Pei-Ching Tsai et al. "Punctuated copy number evolution and clonal stasis in triple-negative breast cancer." *Nature genetics* 48, no. 10 (2016): 1119.

Gilbert, Luke A., Max A. Horlbeck, Britt Adamson, Jacqueline E. Villalta, Yuwen Chen, Evan H. Whitehead, Carla Guimaraes et al. "Genome-scale CRISPR-mediated control of gene repression and activation." *Cell* 159, no. 3 (2014): 647-661.

Gottlieb, Eyal, and Karen H. Vousden. "p53 regulation of metabolic pathways." *Cold Spring Harbor perspectives in biology* 2, no. 4 (2010): a001040.

Gray, Joe, and Brian Druker. "Genomics: the breast cancer landscape." *Nature* 486, no. 7403 (2012): 328.

Greaves, Mel, and Carlo C. Maley. "Clonal evolution in cancer." *Nature* 481, no. 7381 (2012): 306.

Greenblatt, M. S., William P. Bennett, M. Hollstein, and C. C. Harris. "Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis." *Cancer research* 54, no. 18 (1994): 4855-4878.

Greenman, Christopher, Philip Stephens, Raffaella Smith, Gillian L. Dalgliesh, Christopher Hunter, Graham Bignell, Helen Davies et al. "Patterns of somatic mutation in human cancer genomes." *Nature* 446, no. 7132 (2007): 153.

Guha, Tanya, and David Malkin. "Inherited TP53 mutations and the Li–Fraumeni syndrome." *Cold Spring Harbor perspectives in medicine* 7, no. 4 (2017): a026187.

Hafner, Antonina, Martha L. Bulyk, Ashwini Jambhekar, and Galit Lahav. "The multiple mechanisms that regulate p53 activity and cell fate." *Nature reviews. Molecular cell biology*(2019).

Hainaut, Pierre, and Gerd P. Pfeifer. "Somatic TP53 mutations in the era of genome sequencing." *Cold Spring Harbor Perspectives in Medicine* 6, no. 11 (2016): a026179.

Hollestelle, Antoinette, Jord HA Nagel, Marcel Smid, Suzanne Lam, Fons Elstrodt, Marijke Wasielewski, Ser Sue Ng et al. "Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines." *Breast cancer research and treatment* 121, no. 1 (2010): 53-64.
Hollstein, Monica, David Sidransky, Bert Vogelstein, and Curtıs C. Harris. "p53 mutations in human cancers." *Science*253, no. 5015 (1991): 49-53.

Iranzo, Jaime, Iñigo Martincorena, and Eugene V. Koonin. "Cancer-mutation network and the number and specificity of driver mutations." *Proceedings of the National Academy of Sciences* 115, no. 26 (2018): E6010-E6019.

Jiang, Le, Ning Kon, Tongyuan Li, Shang-Jui Wang, Tao Su, Hanina Hibshoosh, Richard Baer, and Wei Gu. "Ferroptosis as a p53-mediated activity during tumour suppression." *Nature*520, no. 7545 (2015): 57.

Joerger, Andreas C., and Alan R. Fersht. "Structural biology of the tumor suppressor p53." *Annu. Rev. Biochem.* 77 (2008): 557-582.

Joerger, Andreas C., and Alan R. Fersht. "The p53 pathway: origins, inactivation in cancer, and emerging therapeutic approaches." *Annual review of biochemistry* 85 (2016): 375-404.

Joerger, Andreas C., and Alan R. Fersht. "The tumor suppressor p53: from structures to drug discovery." *Cold Spring Harbor perspectives in biology* 2, no. 6 (2010): a000919.

Kaiser, Alyssa M., and Laura D. Attardi. "Deconstructing networks of p53-mediated tumor suppression in vivo." *Cell death and differentiation* 25, no. 1 (2018): 93.

Kalimutho, Murugan, Katia Nones, Sriganesh Srihari, Pascal HG Duijf, Nicola Waddell, and Kum Kum Khanna. "Patterns of genomic instability in breast cancer." *Trends in pharmacological sciences* (2019).

Kandioler-Eckersberger, Daniela, Carmen Ludwig, Margarethe Rudas, Sonja Kappel, Elisabeth Janschek, Catharina Wenzel, Hermine Schlagbauer-Wadl et al. "TP53 mutation and p53 overexpression for prediction of response to neoadjuvant treatment in breast cancer patients." *Clinical Cancer Research*6, no. 1 (2000): 50-56.

Kim, Charissa, Ruli Gao, Emi Sei, Rachel Brandt, Johan Hartman, Thomas Hatschek, Nicola Crosetto, Theodoros Foukakis, and Nicholas E. Navin. "Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing." *Cell* 173, no. 4 (2018): 879-893.

Klebanov, Nikolai, Mykyta Artomov, William B. Goggins, Emma Daly, Mark J. Daly, and Hensin Tsao. "Burden of unique and low prevalence somatic mutations correlates with cancer survival." *Scientific reports* 9, no. 1 (2019): 4848.

Korkmaz, Gozde, Rui Lopes, Alejandro P. Ugalde, Ekaterina Nevedomskaya, Ruiqi Han, Ksenia Myacheva, Wilbert Zwart, Ran Elkon, and Reuven Agami. "Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9." *Nature biotechnology* 34, no. 2 (2016): 192.

Kuperwasser, Charlotte, Gregory D. Hurlbut, Frances S. Kittrell, Ellen S. Dickinson, Rudy Laucirica, Daniel Medina, Stephen P. Naber, and D. Joseph Jerry. "Development of spontaneous mammary tumors in BALB/c p53 heterozygous mice: a model for Li-Fraumeni syndrome." *The American journal of pathology* 157, no. 6 (2000): 2151-2159.

Lane, David P., and Lionel V. Crawford. "T antigen is bound to a host protein in SY40-transformed cells." *Nature* 278, no. 5701 (1979): 261.

Laptenko, O., and C. Prives. "Transcriptional regulation by p53: one protein, many possibilities." *Cell death and differentiation* 13, no. 6 (2006): 951.

Laptenko, Oleg, Idit Shiff, Will Freed-Pastor, Andrew Zupnick, Melissa Mattia, Ella Freulich, Inbal Shamir et al. "The p53 C terminus controls site-specific DNA binding and promotes structural changes within the central DNA binding domain." *Molecular cell* 57, no. 6 (2015): 1034-1046.

Levine, Arnold J., and Moshe Oren. "The first 30 years of p53: growing ever more complex." *Nature reviews cancer* 9, no. 10 (2009): 749.

Linzer, Daniel IH, and Arnold J. Levine. "Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells." *Cell* 17, no. 1 (1979): 43-52.

Linzer, Daniel IH, Warren Maltzman, and Arnold J. Levine. "The SV40 A gene product is required for the production of a 54,000 MW cellular tumor antigen." *Virology* 98, no. 2 (1979): 308-318.
Martincorena, Iñigo, and Peter J. Campbell. "Somatic mutation in cancer and normal cells." *Science* 349, no. 6255 (2015): 1483-1489.

Martincorena, Iñigo, Keiran M. Raine, Moritz Gerstung, Kevin J. Dawson, Kerstin Haase, Peter Van Loo, Helen Davies, Michael R. Stratton, and Peter J. Campbell. "Universal patterns of selection in cancer and somatic tissues." *Cell* 171, no. 5 (2017): 1029-1041.

Marusyk, Andriy, Vanessa Almendro, and Kornelia Polyak. "Intra-tumour heterogeneity: a looking glass for cancer?." *Nature Reviews Cancer* 12, no. 5 (2012): 323.

Masciari, Serena, Deborah A. Dillon, Michelle Rath, Mark Robson, Jeffrey N. Weitzel, Judith Balmana, Stephen B. Gruber et al. "Breast cancer phenotype in women with TP53 germline mutations: a Li-Fraumeni syndrome consortium effort." *Breast cancer research and treatment* 133, no. 3 (2012): 1125-1130.

Mayer, Ingrid A., Vandana G. Abramson, Brian D. Lehmann, and Jennifer A. Pietenpol. "New strategies for triple-negative breast cancer—deciphering the heterogeneity." *Clinical cancer research* 20, no. 4 (2014): 782-790.

McGranahan, Nicholas, and Charles Swanton. "Biological and therapeutic impact of intratumor heterogeneity in cancer evolution." *Cancer cell* 27, no. 1 (2015): 15-26.

Mello, Stephano S., and Laura D. Attardi. "Deciphering p53 signaling in tumor suppression." *Current opinion in cell biology*51 (2018): 65-72.

Muller, Patricia AJ, Karen H. Vousden, and Jim C. Norman. "p53 and its mutants in tumor cell migration and invasion." *The Journal of cell biology* 192, no. 2 (2011): 209-218.

Nik-Zainal, Serena, Helen Davies, Johan Staaf, Manasa Ramakrishna, Dominik Glodzik, Xueqing Zou, Inigo Martincorena et al. "Landscape of somatic mutations in 560 breast cancer whole-genome sequences." *Nature* 534, no. 7605 (2016): 47.

Norberg, Torbjörn, Sigrid Klaar, Gunilla Kärf, Hans Nordgren, Lars Holmberg, and Jonas Bergh. "Increased p53 mutation frequency during tumor progression—results from a breast cancer cohort." *Cancer research* 61, no. 22 (2001): 8317-8321.

Nowell, Peter C. "The clonal evolution of tumor cell populations." *Science* 194, no. 4260 (1976): 23-28.

Olive, Kenneth P., David A. Tuveson, Zachary C. Ruhe, Bob Yin, Nicholas A. Willis, Roderick T. Bronson, Denise Crowley, and Tyler Jacks. "Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome." *Cell* 119, no. 6 (2004): 847-860.

Olivier, Magali, Anita Langer, Patrizia Carrieri, Jonas Bergh, Sigrid Klaar, Jorunn Eyfjord, Charles Theillet et al. "The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer." *Clinical cancer research* 12, no. 4 (2006): 1157-1167.

Olivier, Magali, David E. Goldgar, Nayanta Sodha, Hiroko Ohgaki, Paul Kleihues, Pierre Hainaut, and Rosalind A. Eeles. "Li-Fraumeni and related syndromes: correlation between tumor type, family structure, and TP53 genotype." *Cancer research* 63, no. 20 (2003): 6643-6650.

Parada, Luis F., Hartmut Land, Robert A. Weinberg, David Wolf, and Varda Rotter. "Cooperation between gene encoding p53 tumour antigen and ras in cellular transformation." *Nature*312, no. 5995 (1984): 649.

Pereira, Bernard, Suet-Feung Chin, Oscar M. Rueda, Hans-Kristian Moen Vollan, Elena Provenzano, Helen A. Bardwell, Michelle Pugh et al. "The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes." *Nature communications* 7 (2016): 11479.

Petitjean, Audrey, Ewy Mathe, Shunsuke Kato, Chikashi Ishioka, Sean V. Tavtigian, Pierre Hainaut, and Magali Olivier. "Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database." *Human mutation*28, no. 6 (2007): 622-629.

Platt, Randall J., Sidi Chen, Yang Zhou, Michael J. Yim, Lukasz Swiech, Hannah R. Kempton, James E. Dahlman et al. "CRISPR-Cas9 knockin mice for genome editing and cancer modeling." *Cell* 159, no. 2 (2014): 440-455.

Polyak, Kornelia. "Heterogeneity in breast cancer." *The Journal of clinical investigation* 121, no. 10 (2011): 3786-3788.

Pon, Julia R., and Marco A. Marra. "Driver and passenger mutations in cancer." *Annual Review of Pathology: Mechanisms of Disease* 10 (2015): 25-50.

Rheinbay, Esther, Prasanna Parasuraman, Jonna Grimsby, Grace Tiao, Jesse M. Engreitz, Jaegil Kim, Michael S. Lawrence et al. "Recurrent and functional regulatory mutations in breast cancer." *Nature* 547, no. 7661 (2017): 55.

Rivlin, Noa, Ran Brosh, Moshe Oren, and Varda Rotter. "Mutations in the p53 tumor suppressor gene: important milestones at the various steps of tumorigenesis." *Genes & cancer* 2, no. 4 (2011): 466-474.

Rotter, Varda, Owen N. Witte, Robert Coffman, and David Baltimore. "Abelson murine leukemia virus-induced tumors elicit antibodies against a host cell protein, P50." *Journal of virology* 36, no. 2 (1980): 547-555.

Rotter, Varda. "p53, a transformation-related cellular-encoded protein, can be used as a biochemical marker for the detection of primary mouse tumor cells." *Proceedings of the National Academy of Sciences* 80, no. 9 (1983): 2613-2617.

Sarnow, Peter, Ye Shih Ho, Jim Williams, and Arnold J. Levine. "Adenovirus E1b-58kd tumor antigen and SV40 large tumor antigen are physically associated with the same 54 kd cellular protein in transformed cells." *Cell* 28, no. 2 (1982): 387-394.

Scholl, Claudia, and Stefan Fröhling. "Exploiting rare driver mutations for precision cancer medicine." *Current opinion in genetics & development* 54 (2019): 1-6.

Shah, Sohrab P., Andrew Roth, Rodrigo Goya, Arusha Oloumi, Gavin Ha, Yongjun Zhao, Gulisa Turashvili et al. "The clonal and mutational evolution spectrum of primary triple-negative breast cancers." *Nature* 486, no. 7403 (2012): 395.

Shalem, Ophir, Neville E. Sanjana, and Feng Zhang. "High-throughput functional genomics using CRISPR–Cas9." *Nature Reviews Genetics* 16, no. 5 (2015): 299.

Shalem, Ophir, Neville E. Sanjana, Ella Hartenian, Xi Shi, David A. Scott, Tarjei S. Mikkelsen, Dirk Heckl et al. "Genome-scale CRISPR-Cas9 knockout screening in human cells." *Science* 343, no. 6166 (2014): 84-87.

Silwal-Pandit, Laxmi, Hans Kristian Moen Vollan, Suet-Feung Chin, Oscar M. Rueda, Steven McKinney, Tomo Osako, David A. Quigley et al. "TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance." *Clinical Cancer Research* 20, no. 13 (2014): 3569-3580.

Sørlie, Therese, Charles M. Perou, Robert Tibshirani, Turid Aas, Stephanie Geisler, Hilde Johnsen, Trevor Hastie et al. "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications." *Proceedings of the National Academy of Sciences* 98, no. 19 (2001): 10869-10874.

Srivastava, Shiv, Zhiqiang Zou, Kathleen Pirollo, William Blattner, and Esther H. Chang. "Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li–Fraumeni syndrome." *Nature* 348, no. 6303 (1990): 747.

Stephens, Philip J., Patrick S. Tarpey, Helen Davies, Peter Van Loo, Chris Greenman, David C. Wedge, Serena Nik-Zainal et al. "The landscape of cancer genes and mutational processes in breast cancer." *Nature* 486, no. 7403 (2012): 400.

Stratton, Michael R., Peter J. Campbell, and P. Andrew Futreal. "The cancer genome." *Nature* 458, no. 7239 (2009): 719.

Sullivan, Kelly D., Matthew D. Galbraith, Zdenek Andrysik, and Joaquin M. Espinosa. "Mechanisms of transcriptional regulation by p53." *Cell death and differentiation* 25, no. 1 (2018): 133.

Sun, Ruping, Zheng Hu, and Christina Curtis. "Big Bang tumor growth and clonal evolution." *Cold Spring Harbor perspectives in medicine* 8, no. 5 (2018): a028381.

Tidow, Henning, Roberto Melero, Efstratios Mylonas, Stefan MV Freund, J. Guenter Grossmann, José María Carazo, Dmitri I. Svergun, Mikel Valle, and Alan R. Fersht. "Quaternary structures of tumor suppressor p53 and a specific p53–DNA complex." *Proceedings of the National Academy of Sciences*104, no. 30 (2007): 12324-12329.

Turajlic, Samra, Andrea Sottoriva, Trevor Graham, and Charles Swanton. "Resolving genetic heterogeneity in cancer." *Nature Reviews Genetics* (2019): 1.

Vogelstein, Bert, and Kenneth W. Kinzler. "The multistep nature of cancer." *Trends in genetics* 9, no. 4 (1993): 138-141.

Wang, Yong, Jill Waters, Marco L. Leung, Anna Unruh, Whijae Roh, Xiuqing Shi, Ken Chen et al. "Clonal evolution in breast cancer revealed by single nucleus genome sequencing." *Nature* 512, no. 7513 (2014): 155.

Wells, Mark, Henning Tidow, Trevor J. Rutherford, Phineus Markwick, Malene Ringkjobing Jensen, Efstratios Mylonas, Dmitri I. Svergun, Martin Blackledge, and Alan R. Fersht. "Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain." *Proceedings of the National academy of Sciences* 105, no. 15 (2008): 5762-5767.

Whyte, Peter, Karen J. Buchkovich, Jonathan M. Horowitz, Stephen H. Friend, Margaret Raybuck, Robert A. Weinberg, and Ed Harlow. "Association between an oncogene and an anti-oncogene: the adenovirus E1A proteins bind to the retinoblastoma gene product." *Nature* 334, no. 6178 (1988): 124.

Williams, Marc J., Andrea Sottoriva, and Trevor A. Graham. "Measuring Clonal Evolution in Cancer with Genomics." *Annual review of genomics and human genetics* 20 (2019).

Wolf, D. A. V. I. D., and V. A. R. D. A. Rotter. "Inactivation of p53 gene expression by an insertion of Moloney murine leukemia virus-like DNA sequences." *Molecular and cellular biology* 4, no. 7 (1984): 1402-1410.

Wolf, David, and Varda Rotter. "Major deletions in the gene encoding the p53 tumor antigen cause lack of p53 expression in HL-60 cells." *Proceedings of the National Academy of Sciences* 82, no. 3 (1985): 790-794.

Wolf, David, Nicholas Harris, and Varda Rotter. "Reconstitution of p53 expression in a nonproducer Ab-MuLV-transformed cell line by transfection of a functional p53 gene." *Cell* 38, no. 1 (1984): 119-126.

Wood, Laura D., D. Williams Parsons, Siân Jones, Jimmy Lin, Tobias Sjöblom, Rebecca J. Leary, Dong Shen et al. "The genomic landscapes of human breast and colorectal cancers." *Science* 318, no. 5853 (2007): 1108-1113.

Yates, Lucy R., Moritz Gerstung, Stian Knappskog, Christine Desmedt, Gunes Gundem, Peter Van Loo, Turid Aas et al. "Subclonal diversification of primary breast cancer revealed by multiregion sequencing." *Nature medicine* 21, no. 7 (2015): 751.

Yates, Lucy R., Stian Knappskog, David Wedge, James HR Farmery, Santiago Gonzalez, Inigo Martincorena, Ludmil B. Alexandrov et al. "Genomic evolution of breast cancer metastasis and relapse." *Cancer Cell* 32, no. 2 (2017): 169-184.

Zardavas, Dimitrios, Alexandre Irrthum, Charles Swanton, and Martine Piccart. "Clinical management of breast cancer heterogeneity." *Nature reviews Clinical oncology* 12, no. 7 (2015): 381.

Zhu, Shiyou, Wei Li, Jingze Liu, Chen-Hao Chen, Qi Liao, Ping Xu, Han Xu et al. "Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library." *Nature biotechnology* 34, no. 12 (2016): 1279.

Zilfou, Jack T., and Scott W. Lowe. "Tumor suppressive functions of p53." *Cold Spring Harbor perspectives in biology*1, no. 5 (2009): a001883.

CHAPTER 2

DISTINCT NEOMORPHIC MOLECULAR AND CELLULAR FUNCTIONS OF DIFFERENT MISSENSE MUTANT P53

Introduction

The tumor suppressor *TP53* is the most frequently mutated gene in human cancers, and, in different cancer types, mutations in *TP53* occur at different stages of malignant transformation and contribute to tumor initiation, progression or metastatic spread (Rivlin 2011). Unlike other classical tumor suppressor genes like *RB, APC* and *BRCA1* that are typically deleted or truncated in cancer, the majority of *TP53* mutations occur as missense mutations, concentrated within the exons 4 to 9 that encode the DNA binding domain of the transcription factor. p53 functions as a homo-tetramer and this provides the mutant p53 protein with the privilege of dominant negative action over the wildtype allele when the both alleles co-exist in a heterozygous condition. However, loss of heterozygosity (LOH) follows *TP53* mutations during cancer progression where the remaining allele is deleted or mutated. LOH is commonly seen in tumor suppressors. Following LOH, mutant p53 gains additional function, gain-of-function (GOF), oncogene like activities than simply loss of tumor suppressor function.

The timing of p53 mutation occurrence during tumorigenesis is variable for different cancer types and suggests the function of p53 at different stages of cancer development. p53 mutations appear at a later stage in colorectal carcinomas (CRCs), prostate and bladder cancer. On the other hand, mutant p53 is found in human pre-malignant breast lesion, ductal carcinoma in-situ (DCIS) and in Barret esophagus patients that predisposes them to esophageal adenocarcinoma.

Though *TP53* mutations are most frequent genetic aberrations found in different cancer types, there is limited knowledge about the effect of the different p53 mutants on cancer-related

phenotypes and the underlying mechanisms. Since more than 75% of the mutations in *TP53* are the missense mutations that cluster in the DNA binding domain of this transcription factor, a comprehensive study on whether and how these different missense mutations have heterogeneous influences on cellular phenotypes and molecular function of p53 proteins is needed. It can shed light on understanding the phenotypic heterogeneity observed in cancer tumors and finding personalized therapeutic options particularly for TNBC patients as nearly 88% of these tumors harbor mutation in *TP53*. Therefore, as the first step, we aimed to generate a panel of cell lines that mimic triple-negative tumors with a wide range of different p53 missense mutations and investigate their molecular and cellular characteristics.

Aim

Phenotypic characterization of ten most prevalent p53 missense mutant proteins found in breast cancer tumors expressed in non-transformed mammary epithelial cells and molecular profiling to elucidate their phenotypic heterogeneity.

Approach

Model system of normal breast epithelial cells to study the effect of mutant p53

MCF 10 breast epithelial cells were derived from a pre-menopausal woman with fibrocystic disease that spontaneously immortalized into MCF 10A cells capable of attachment and growth in DMEM-F12 media (Tait 1990). The MCF 10A cells are the most widely used non-transformed mammary epithelial cell line, which are ERα negative and behave mostly like basal cells though they carry luminal markers. The cell line carries wild type *TP53*, while the p16 and p14 genes are absent. In 3D culture, they form an acinus like spheroid with a hollow lumen and polarized single

layer of cells with basement membrane on the outside. Therefore, it has been a useful tool to study mammary gland development, effect of microenvironment and genetic alterations on mammary cell transformation.

TP53 missense mutations are the most frequent in the basal-like or the triple-negative breast cancer subtype. Hence, for studying the phenotypic effect of these mutations, the receptor negative (*i.e*., no ERα, no PR and no HER2 amplification) MCF 10A cell-line was the natural choice as opposed to the MCF 12A cells, which are ERα positive non-tumorigenic mammary epithelial cells with sensitivity to estrogen (Marchese 2012). In another breast epithelial cell line of the immortalized Human Mammary Epithelial Cells (HMECs) with human (h) TERT (catalytic subunit of telomerase) resulted in activation and overexpression of *c-myc* in the cells (Wang 2000). To study the functional effect of the different missense mutant p53 occurring in breast cancer, we over-expressed the mutant p53 proteins in the normal-like MCF 10A and the HMEC-hTERT cells. However, the missense mutant p53 over-expressing HMEC-hTERT cells provided very inconsistent results in the phenotype assays while the p53 mutant-expressing MCF 10A cells displayed consistent phenotypes. Therefore, we chose MCF 10A as the cell-based model system for this study.

The phenotype assays

The transformed cancer cells acquire certain molecular and biochemical capabilities that are shared by most cancer cells, and these new traits dictate malignant growth. The cellular traits include sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis and activation of cell invasion and metastasis (Hanahan 2011). These acquired capabilities comprise the ‘hallmarks of cancer’, and mutation in *TP53* directly affects cell viability and proliferation potential, apoptosis (programmed cell death) resistance, cell migration and invasion among the hallmarks. Therefore, we chose to characterize 1) cell survival and proliferation in absence of growth factors 2) resistance to

apoptosis and anoikis 3) cell migration and invasion and 4) 3D morphology of mammospheres. Furthermore, well established assays were available for measuring these phenotypes, and the assays were easy to adapt to a high-throughput format.

Molecular profiling of the p53 mutant MCF 10A cell lines

The p53 is a transcription factor that binds to DNA elements with specific sequence motifs and modulate the transcription of target genes. The missense mutations on the DNA binding domain affect the DNA binding capacity and preference of p53. Therefore, we included the ChIP seq and RNA seq studies in addition to the phenotype assays to understand the molecular mechanisms giving rise to the phenotypic differences. ChIP seq would highlight the DNA binding targets of the p53 mutants that could be correlated with up and down regulation of the corresponding genes measured by RNA seq. Lastly, integrated bioinformatics analysis can correlate the changes in overall gene expression profiles with the phenotypes induced by different p53 mutants and infer the dysregulated pathways underlying the heterogeneous phenotypes.

Method

Cell culture

MCF 10A cells were cultured at 37˚C in a 5% CO_2 humidified incubator. DMEM/F12 media (ThermoFisher #11320-082) was supplemented with 5% Horse Serum (ThermoFisher #16050-122), hEGF (10ng/ml, ThermoFisher #PHG0311), hydrocortisone (0.5µg/ml, Sigma #H0888), cholera toxin (100ng/ml, Sigma #C8052), and Insulin (10µg/ml, Sigma #I9278). Cells were routinely passaged with trypsin (0.25% in HBSS with 0.2g/ml EDTA, GE Healthcare

#SH30042.01) at 80-90% confluence. The cells were checked for mycoplasma infection at regular intervals as it is a common cell culture contaminant.

Production of MCF 10A clones

The MCF 10A cells over-expressing p53 mutant proteins were made using vectors procured from DNASU (The center for personalized diagnostics at the Biodesign Institute, ASU). The *TP53* missense mutant genes from these vectors were cloned into the pLenti4/V5-DEST Gateway vector (ThermoFisher Scientific #V49810). Plasmids were expanded in 500ml *E. coli* cultures and then purified using Machery-Nagel DNA Maxi-Preps. The stable cell-lines were then prepared through lentiviral transduction.

Western blot

Cells were grown to 80% confluency in six well plates (Greiner #5665-7160) and lysed using RIPA lysis buffer that contains 50mM Tris, 150mM NaCl, 1% IGEPAL CA-630 (Sigma #I9996), 0.1% sodium azide (Sigma #S8032), cOmplete mini protease inhibitor (Sigma #04693124001), 200 µM sodium fluoride (Sigma #S6776) and 200 µM of sodium orthovanadate (Sigma #450243). Cells were scraped, incubated on ice for 10 minutes and then centrifuged at 10,000 rpm. Lysates were quantified using the Pierce BCA kit (#23225). Samples were run on 4-20% TGX gels from BioRad (#567-1093) and blotted onto 0.45µm PVDF membrane (GE Healthcare #10600023) using the BioRad semi-dry transfer system. Primary antibody for p53 protein from Sigma (#P6874) diluted to 1:200 and secondary HRP-linked anti-mouse antibody (Cell Signaling Technologies #7076) diluted to 1:3000 were used to visualize p53 bands. The p53 mutant proteins were seen using the V5 tag Ab (CST, Catalogue no. 13202S). The western blots were imaged on the Fluorchem FC2 from AlphaInnotech.

Assay for cell viability

Cells were parallelly plated in clear CellBIND 96-well plates (Corning #3300) for pictures and opaque white plates for chemiluminescent analysis (Perkin-Elmer #6005680) and allowed to grow for 24 hours in normal growth media before treatment. The following day, media was replaced with the restricted media with or without growth factors serum and EGF (+S+E, -S-E, +S-E and -S+E). Human EGF (10 ng/ml, ThermoFisher #PHG0311) and horse serum (5% v/v, ThermoFisher #16050-122) were excluded individually or together from normal growth media and added to the cells. Cells were then allowed to grow for 72 hours and cell viability was analyzed by Cell Titer Glo (Promega # PAG7571) on the Envision plate reader (Perkin-Elmer). Results were calculated with respect to untreated cells, then $\log_2$ transformed.

Apoptosis assay

MCF 10A p53 mutant cells (4000 cells/well) were plated into both 96-well CellBIND clear plates (Corning #3300) for observation and Perkin-Elmer white opaque plates (#6005680). On the following day, cells were treated with 1µM Doxorubicin (Calbiochem #324380) using the Biomek NX liquid handler (Beckman Coulter) and incubated for 24 hours. For evaluating apoptosis induction, cells were treated with Caspase-Glo 3/7 (Promega #G8091) per supplier protocol and read in the Envision plate reader (Perkin-Elmer). Results were reported as $\log_2$ fold-change relative to the MCF 10A *TP53* WT-OE cells.

Cell migration and invasion assay

MCF 10A cell-lines (70-80% confluency) were trypsinized and counted using trypan blue dye. The invasion chambers were taken out of -20˚C, and 200 µl of serum free media was added to the inserts, incubated at 37˚C for at least 1hr to warm the Matrigel coat. 2.5 x 10^5 cells (in 200ul) in serum free media were seeded in the upper well of the Matrigel coated (Corning cat no. 354480

for invasion assay) or non-coated (Corning cat no. 354569, migration assay) cell culture insert. The media with serum (750ul) as the chemoattractant was placed in the lower well. The plates were incubated for 22 hrs. The migratory or invasive cells were first fixed with 4% paraformaldehyde for 2 min at RT, permeabilized by adding 500ul of 100% cold methanol (2 min at RT) to each well and stained with 0.5% crystal violet. A media-moistened cotton swab was used to remove any cells that did not migrate or invade through the membrane. Cells were imaged under a microscope at 4X and 10X magnification. Cells were counted in CellProfiler from different areas of the membrane to get a more accurate representation of cell count.

Anoikis assay

MCF 10A p53 mutant and control cells lines were plated into 96-well round bottom (low binding) plates at 150 cells/well and allowed to grow in regular growth media for 7 days. After 7 days, Vybrant Violet Dye (ThermoFisher Scientific #V35003) was added to the cell aggregates at a final concentration of 0.5µM and incubated at 37˚C for a minimum of 30 min. Then Ethidium Homodimer (Biotium #40014) was added to the wells at a final concentration of 0.3µM directly before analysis. The MetaXpress High Content Image Acquisition platform (Molecular Devices) was used to capture images from all the wells of the plate. A total of eight wells for each sample were captured and only the images of the entire cell mass was used for image analysis. Vybrant Violet and the Ethidium homodimer were captured in the Cy3 and Cy5 channels respectively. ImageXpress image analysis software (Molecular Devices) was used to identify and count the cells which were stained with either Vybrant Violet (blue) or Ethidium Homodimer (red). Data was reported as a ratio of live cells to dead cells (Cy3:Cy5) and $\log_2$ transformed.

3D Culture and Analysis

Before adding cells to wells, 96-well black µClear plates were pre-coated with 100µl of Matrigel (Corning #35428) and allowed to gel for 30 min. Cells were added on top of the Matrigel and

allowed to grow for nine days, with two media changes in the interim. Cells were then fixed with 4% paraformaldehyde (Alfa-Aeser # AA43368-9M) for 15 minutes and were washed with 100mM Glycine-PBS solution to neutralize the paraformaldehyde and then permeabilized with 0.5% Triton X-100 (Sigma-Aldrich #X100) for 15 min at 37˚C. Permeabilized cells were blocked with 10% goat serum (ThermoFisher Scientific #PCN5000) and 20ug/ml goat α-mouse IgG (Jackson ImmunoResearch #115-006-006) for at least 1hr. The spheroids were washed with PBS and plates stored at 4˚C until use. Staining of cells were done following the procedure previously described (Debnath 2003). Analysis was performed by employing custom Cell Profiler pipelines. For identification of spheroids, the β-Catenin staining was used as a mask. Intensity measurements were then based on the average pixel intensity over the object mask for each target protein. To analyze laminin localization, relative bins (concentric rings) were created around the center of each object. The average pixel intensity per ring was recorded and graphed, giving intensity measurements relative to the center of each spheroid. For nuclear localization (clearing of lumens) the same technique of using bins was employed, analyzing nuclear intensity per bin.

RNA seq

Total RNA extracted from the cells (Qiagen catalogue no. 74134) were used to prepare cDNA using Nugen's Ovation RNA-Seq System via single primer isothermal amplification (Catalogue # 7102-A01) automated on the Apollo 324 liquid handler from Wafergen. cDNA was sheared to approximately 300 bp fragments using the Covaris M220 ultrasonicator. Libraries were generated using Kapa Biosystem's library preparation kit (KK8201). Fragments were end-repaired and A-tailed, individual indexes and adapters (Bioo, catalogue #520999) were ligated on each separate sample. The adapter ligated molecules were cleaned using AMPure beads (Agencourt Bioscience/Beckman Coulter, A63883), and amplified with Kapa's HIFI enzyme. The library was then analyzed on an Agilent Bioanalyzer, and quantified by qPCR (KAPA Library Quantification

Kit, KK4835) before multiplex pooling and sequencing a 2x75 PE flow cell on the NextSeq500 platform (Illumina) at the ASU's CLAS Genomics Core facility.

Raw sequencing read data quality were analyzed using FastQC (v0.10.1). STAR (V020201) (Dobin 2013) was used to align reads to the human genome (GRCh38.p92/hg38) to counts. Duplication rate was obtained, and duplicated reads were marked by PICARD (PICARD 2.18.3). Duplication plots were generated using R package dupRadar (Sayols 2016). As high duplication rates were observed (85.81% - 97.24%), gene counts after removal of duplicated reads were used as a reference to check whether the expression levels were overestimated due to PCR duplication. Next, because mitochondrial RNA occupied a large but variable portion of RNA pool in every sample (40 to 80% of all sequence reads), to avoid the bias in calculating the total read-normalized transcript abundance, mitochondrial RNA data was dropped. Finally, sequencing counts were normalized to TPM (transcripts per million reads).

To validate that the duplicated reads did not systematically impact the quantification of transcript levels, we repeated RNA-Seq on a subset of Y234C, Y220C, R273C, R273H and WT-OE cell lines and compare the overall expression profiles. Despite the much lower duplication rates (<60%) in the second RNA Seq, the Pearson's correlation coefficient of RNA expression levels between the runs ranged from 0.7 to 0.82.

ChIP seq

All the MCF 10A cell lines were fixed with 1% formaldehyde in PBS and incubated for 10 min at RT in the petri-dish. Formaldehyde was quenched with glycine (125mM final conc.) for 5 min at RT. Cells were rinsed with ice cold PBS twice and scraped with 2ml of cold PBS with protease inhibitor (Sigma Aldrich Cat no. 04693124001). Cells were washed two more times with cold PBS with protease inhibitor and pellets from 20 million cells were snap frozen and stored at -80^{o}C. 1ml of ChIP lysis buffer (1% SDS,10mM EDTA, 50mM Tris-HCl pH8.1) with protease inhibitors was

added and the cells were sonicated in the Covaris M220 ultrasonicator for 8 min. The time and setting for sonication would vary with the cell line and concentration. The lysates were spun at maximum speed at 4^{o}C for 10 min and the DNA concentration was checked. The supernatant was transferred into fresh tube. 25-50ul of the lysate was incubated at 65^{o}C overnight with 0.1M NaCl and run on a 2% agarose gel to check the sonication efficiency. You should get an average fragmentation of ~500 bp. For IP, 30ul of beads (Thermo Catalogue no. 11203D) were first washed with 1ml of cold BSA (5mg/ml) in PBS at RT (3 times). Then V5 antibody (CST Catalogue no. 13202S) was added to the beads in 1ml of PBS+ BSA. The beads with the antibody were incubated for 4 to 6 hrs at 4^{o}C. The IP was performed on 50ug of total DNA and the samples were diluted accordingly in the dilution buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH 8.0). The diluted chromatin was precleared with 15ul of washed beads for 1hr at 4^{o}C. 5% of this precleared lysate was kept as input control. PBS/BSA was aspirated from the antibody coated beads and precleared lysate was added to incubate O/N at 4^{o}C on a rotating wheel. Next day beads were collected using the magnetic concentrator and washed 6 times with ChIP RIPA buffer (50mM HEPES, 1mM EDTA, 0.7% Na Deoxycholate, 1% NP-40, 0.5M LiCl pH 7.6) at RT on a rotating platform for 10 min between every wash. Then washed twice with 1X TE (pH7.6) at RT. 100 ul of Elution buffer (1% SDS, 0.1M NaHCO3, 0.1M NaCl) was added to the beads and input samples (make up volume to 100ul). The beads were vortexed in this solution every few minutes for 30 minutes in total at RT and incubated at 65^{o}C O/N for 12-14 hrs. Next day 1ul of proteinase K (10mg/ml) was added and incubated at 42^{o}C for 2 hrs. The IP DNA was purified with QIAquick PCR purification kit.

The sequencing read counts varied from 4.3 X 10^{7} to 5.9 X 10^{7}. Quality of the reads were analyzed using FastQC (v0.10.1). Paired-ended reads were mapped to the reference human genome (GRCh38.p92/hg38) end to end using Bowtie2 (Bowtie2 2.1.0) (Langmead 2012). Non-primary alignment, unmapped reads were removed by Samtools (Samtools v1.7) (Li 2009). Duplication rates were calculated by PICARD (PICARD 2.18.3) (duplication rate ranges from 18% to 88%). MACS2 (macs2 2.1.2.20181017) was used for peak calling using BAMPE mode under

the q-value cutoff 0.05 (Zhang 2008). Peaks were annotated under Ensembl genome annotation (GRCh38.p92/hg38) using Homer toolkit (Homer v4.9.1) (Heinz 2010), which was also been utilized for motif finding (Heinz 2010). Regions located between -100 and 2500 bp upstream of the nearest TSS were defined as the promoter region. Known and de-novo motifs were called by Homer toolkit. De-novo motifs comparison was done by Tomtom in MEME Suite (MEME Suite 5.0.1) (Timothy 2009).

Pathway analysis

Partial Least Squares Regression models were built for each of the KEGG pathways under each phenotype using scikit-learn, a machine learning Python library (Pedregosawe 2011). Gene expression values of the cell lines were used as training vectors and the phenotype scores of the cell lines under each phenotype were used as targeting vectors. Leave one out cross validation was implied to avoid overfit. Mean value for R^2 and mean squared of error (MSE) were calculated in the test set for evaluating the performance of the models and ranking the pathways under a particular phenotype. Visualized pathway maps were generated by Cytoscape (Shannon 2003) and module KEGGscape (Nishida 2014).

GSEA

Gene Set Enrichment Analysis (GSEA) was carried out using gseapy, a Python library (Subramanian 2005). The gene sets were obtained from ChIP Seq data using lists of gene that were targeted at promoter region. The ‘jump’ score of each gene was calculated using RNA expression score (TPM) minus the mean TPM of this gene in all the cell lines. Then genes were divided into two groups by the cut-off of 0, and GSEA was carried out on two groups in each cell line separately.

Flow cytometry

MCF 10A p53 mutant cells were grown in 6-well plates (Greiner #5665-7160) for 24 hours and then treated with 1µM Doxorubicin (VWR #80058-048) for 24 hours. Staurosporine (Cell Signaling #9953) at 0.1µM was the positive control for the experiments. Cells were washed in PBS and resuspended in annexin binding (ThermoFisher # V13246) buffer and counted. Cells were aliquoted to 1,000,000 cells/ml in 100µl for annexin and PI binding. After transfer to 1.5ml microcentrifuge tubes, cells were reconstituted in 100 µl annexin binding buffer and then 5ul of annexin (ThermoFisher #A13199) was added and incubated at 4˚C for 15 minutes. After the incubation, 400µl of annexin binding buffer was added followed by propidium iodide (PI; ThermoFisher #P3566) to a final concentration of 1 µg/ml. The samples were kept on ice while analyzing on the Attune flow cytometer from Life Technologies. Compensation controls were run to apply compensation settings, as well as unstained controls to set FSC and SSC parameters. Gates were set on the negative controls.

Cell migration and invasion in 96-well plates

Matrigel (BD Matrigel 354230, stored at -20^{o}C) was thawed while keeping in ice in a 4^{o}C refrigerator overnight. It was diluted to 3mg/ml in ice-cold serum free media. The collagen coated 96-well plate (Platypus tech Cat no. CMACC5.101) was removed from 4^{o}C and allowed to equilibrate to RT. The underside of the plate was checked to ensure that the seeding stoppers were firmly sealed against the bottom of the plate. 100uL of suspended cells were added into each well (control & test wells) without disturbing the seeding stopper, lightly tapping the plate to evenly distribute the cells in each well. The seeded plate with stoppers was incubated at 37^{o}C & 5% CO_2 for 6-8 hours to permit cell attachment. Using the stopper tool seeding stoppers were removed. Reference or the control wells remained with the stoppers till the end of the assay. Media was removed with a pipette and the well gently washed with 100uL of serum free media to remove any unattached cells. 50uL of the Matrigel overlay was added to each well. The plate was

incubated at 37°C and 5% CO_2 for 1hr to permit the Matrigel polymerization. 100uL of serum media was a put on top of the 3D matrix. The plate was incubated at 37°C for 18-20 hours to permit cell invasion. For end-point analysis cell seeding stoppers were removed from the reference/control wells, Matrigel overlay was added and allowed to form the gel. The cells were stained with 5uM cell tracker green (100ul/ well, 45 min, 37°C) and 0.5uM vybrant violet dye (100ul/ well, 30 min, 37°C). Wells were washed with 1X PBS and fixed using 3.7% paraformaldehyde in 1X PBS for 15 min at RT. Images were taken with ImageXpress at different planes on the Z axis.

Immunofluorescence

This protocol was used for staining cells in a 96-well (VWR #82050-748) format and a shaker was used during antibody incubations to obtain uniform staining. The cells were washed with 1XPBS (100ul, 2 times) and fixed with 75µl of 4% paraformaldehyde (VWR #AA43368-9M) per well (15 min at RT). Cells were washed with PBS and 75µl of Blocking Buffer (5% Goat Serum Life Technologies #PCN5000 with 0.32% Triton X-100 Sigma #T8787 in 1X PBS) added per well (1hr at RT). After PBS wash, 50ul/well of primary antibody (at recommended dilutions in Ab protocol) in the Antibody Buffer (1% BSA and 0.3% Triton X-100 in 1X PBS) was added and incubated at RT, shaking gently at 350 RPM in an orbital shaker for 1-2 hours for same day processing or O/N at 4°C while shaking. Cells were washed with PBS and 50ul/well of secondary antibody in Antibody Buffer (at recommended dilutions in Ab protocol) was added with incubation at RT, shaking at 350 RPM in an orbital shaker for 1-2 hours. After final wash in 1x PBS, plates were wrapped in aluminum foil and stored at 4°C with 100µl 0.05% Sodium Azide (NaN_3 Sigma #S8032) in PBS.

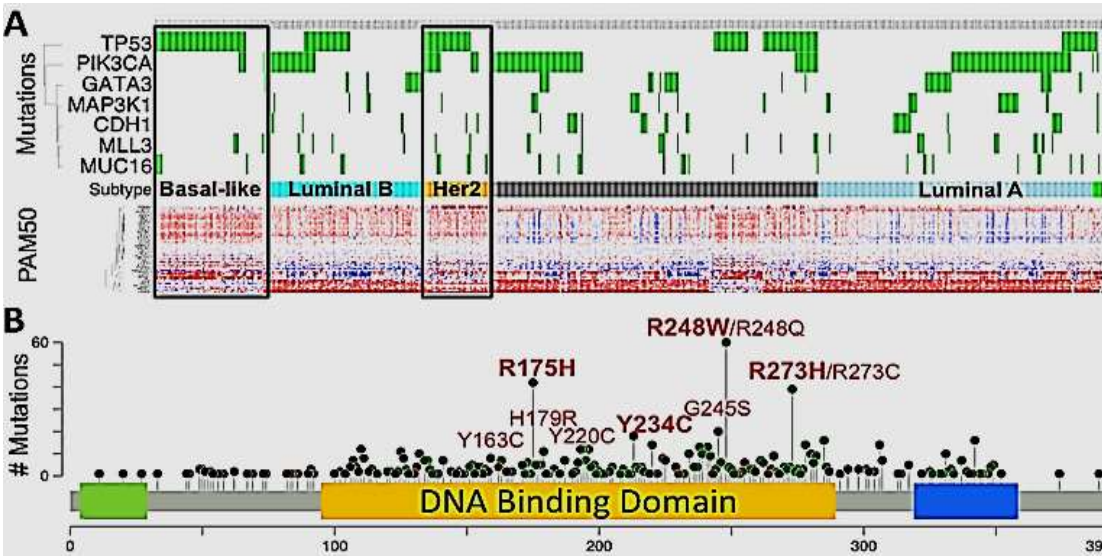
Results

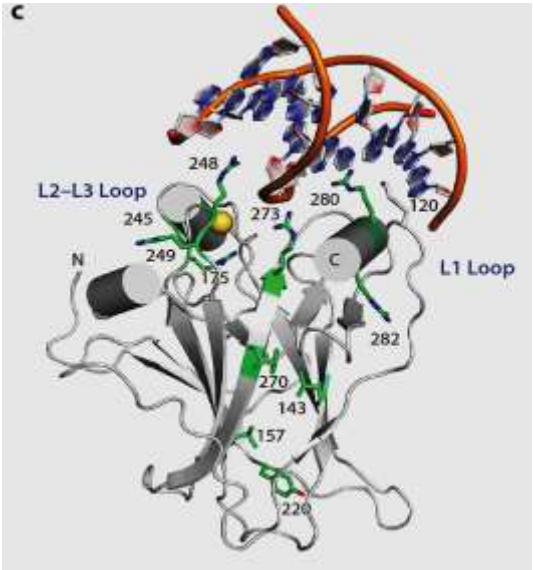
Generation of MCF 10A cell lines expressing 10 different clinically important mutant p53 proteins

More than 100 different missense mutations occur across the *TP53* gene, and most of them concentrate within the DBD of the protein, thus influencing the DNA binding of p53. Although a limited amount of study on a few p53 missense mutants showed potential neo-morphic gain-of-functions of the individual mutants (Freed-Pastor 2012), it was still not entirely clear if the different missense mutant p53 proteins are functionally equal or unique. Therefore, to address the question, we designed a cell line-based study to characterize the influence of a broad spectrum of *TP53* missense mutations on molecular profiles and cellular phenotypes. To focus our study on clinically relevant mutations, we first selected the ten most prevalent *TP53* missense mutations found in breast cancer patients by using the TCGA data (Fig 2.1B). Among the ten mutations, four mutations (R248Q, R248W, R273C, and R273H) are at the residues in direct contact with DNA (referred as “DNA contact mutations”), whereas the other six mutations (G245S, H179R, R175H, Y163C, Y220C, and Y234C) occur in other areas of the DNA binding domain thus affecting the structure of the domain (“Structural mutants”). Thus, in addition to the differences among the individual mutations, it was of interest to see whether the two types of mutations affect distinctively the cellular phenotypes, DNA binding properties, and overall gene expression profiles. As described in the Introduction, we utilized the normal mammary epithelial cell line MCF 10A lacking ER and PR expression as a TNBC-like model system to generate cell lines stably expressing mutant p53 proteins. It should be noted that the cell lines express the endogenous wildtype p53 protein, modeling the heterozygous *TP53* mutations.

The missense mutant *TP53* genes were cloned in the pLenti4/V5-DEST Gateway vector, packaged into lentiviruses, transduced to MCF 10A cells, and selected for the over-expression of the p53 mutant proteins. Western blots showed consistently higher p53 mutant protein levels with

respect to MCF 10A cells with endogenous wildtype p53 protein. However, more importantly, the mutant protein levels were comparable to those of a panel of breast cancer cells harboring missense mutations in endgenous *TP53* gene such as HCC70 (*TP53*^{R248Q}), MDA-MB-231 (*TP53*^{R280K}), MDA-MB-468 (*TP53*^{R273H}), AU565 (*TP53*^{R175H}), and SK-BR-3 (*TP53* ^{R175H}) (Fig 2.2).





Structural	DNA contact
G245S	R248Q
H179R	R248W
R175H	R273C
Y163C	R273H
Y220C	
Y234C	

Figure 2.1. **The TCGA data on *TP53* mutations in breast cancer**. A. The frequency of *TP53* mutations is the highest in the basal-like subtype. The mutation type includes missense and truncation mutations. B. The highly frequent missense mutations concentrate at the DNA binding domain (DBD) of transcription factor p53 and influence DNA binding. C. The amino acid positions of the DNA contact and the structural mutants of p53. Image adapted from (Joerger 2016).

Characterization of cancer hallmark phenotypes of mutant p53 protein expressing MCF 10A cells

We developed high-throughput functional assays to investigate how the different missense mutant p53 proteins affect the cell phenotypes. These quantitative assays assessed several hallmarks of cancer including 1) cell viability in the complete absence of essential growth factors 2) resistance to apoptosis 3) cell migration and invasion 4) anoikis resistance and 5) changes in the 3D acinar morphology and cell polarity. For an unbiased analysis of the phenotypic outcomes and to have a uniform genetic background for comparison among the mutants, we also developed MCF 10A cell line over-expressing the wildtype p53 protein (MCF 10A $p53^{OE}$ or WT^{OE}). All the phenotype data was normalized with respect to the phenotype observed in MCF 10A $p53^{OE}$ cells. MCF 10A $p53^{KD}$ (endogenous p53 knocked down with shRNA) cells were also included in the phenotypic assays to underscore the relevance of dominant negative activity of the missense mutant p53 proteins over the endogenous wildtype p53. All the phenotypes were scored by the log2-transformed fold changes over the wildtype p53 values for cross-comparisons and for plotting.

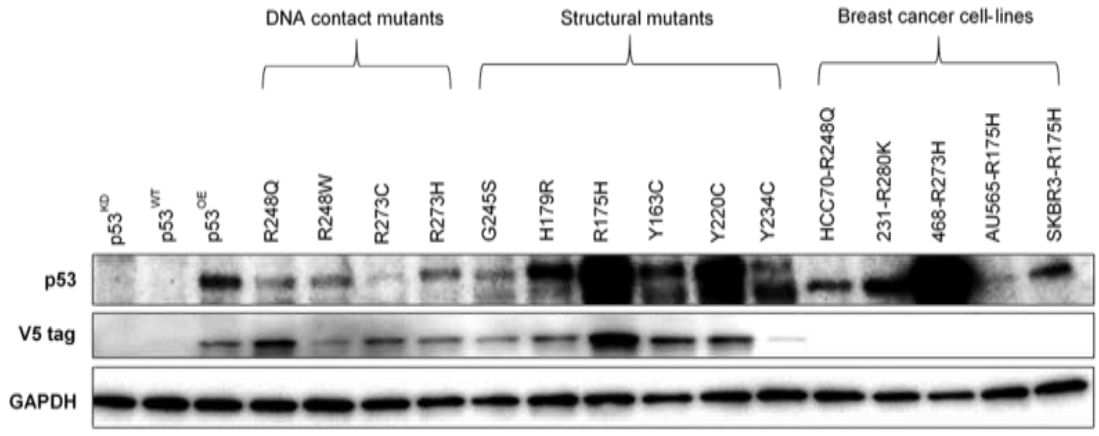


Figure 2.2. **Western blot of MCF 10A p53 mutant cell lines confirming the stable overexpression of mutant p53 proteins.** The total p53 protein level in the cells is comparable to the other breast cancer cell lines harboring a mutation in endogenous *TP53* (HCC70, MDA-MB 231, MDA-MB 468, AU565 and SKBR3). The mutant p53 proteins expressed in MCF 10A cells have a C-term V5 tag. Band for V5 tag is seen for all the mutants and for wildtype p53 over-expressing (OE) cells.

Survival and proliferation of cells in absence of growth factors

A fundamental characteristic of cancer cells is the ability to survive and sustain high tumor proliferation rate even in the absence of growth factors (Hanahan 2011). To find out whether missense p53 mutations by themselves are sufficient to induce self-sustained proliferation without presence of external growth factors and whether different p53 mutants affect the process at different degree, we tested the proliferation and viability of the ten MCF 10A mutant p53 expressing cells in the presence or absence of growth factors.

The MCF 10A cells expressing mutant p53 proteins were kept in the complete DMEM-F12 media supplemented with EGF (epidermal growth factor), insulin, hydrocortisone, cholera toxin and 5% horse serum, and switched to the minimal DMEM-F12 media with or without growth factors (*i.e*., serum and EGF). We evaluated the viability of all the ten mutant p53 cells under different growth conditions, with or without serum and EGF (-Serum+EGF, +Serum-EGF, and –Serum-EGF). After 72 hours, the CellTiter-glo assay was used to determine the number of viable cells in culture after incubation in the different growth conditions. The luminescent signal recorded was proportional to the amount of ATP present that directly corresponded to the number of live cells.

In general, the DNA contact mutants R248Q, R248W, R273C and R273H had higher viability than the structural mutants after 72 hrs of incubation in complete absence of serum and EGF (-S-E) (Fig. 2.3A). R248Q was the most viable mutant with a viability significantly greater than the MCF10A p53 WT, WT-OE, as well as the $p53^{KD}$ cells indicating the neo-morphic activity of the mutant over the loss-of-function effects. R175H and Y234C were the only structural mutants that showed viabilities comparable to that of the MCF10A $p53^{KD}$ cells under the growth factor starved condition. In contrast, withdrawal of EGF from the growth media significantly decreased the viability of all the cell lines and had a profound effect on viability of the mutant G245S (Fig 2.3C).

To measure proliferation, we counted the cells under the same experimental conditions with DAPI staining, and high nuclei counts for the R248Q and R175H cells in the absence of serum and EGF were observed, demonstrating an increase capability in self-sustained survival and proliferation without external stimuli (Fig 2.3B). In agreement, mouse embryonic fibroblasts (MEFs) extracted from the genetically engineered $p53^{R172H/+}$ (equivalent to human R175H) mice were more proliferative with a larger fraction of cells in the S-phase compared to $p53^{+/-}$ MEF (Olive 2004).

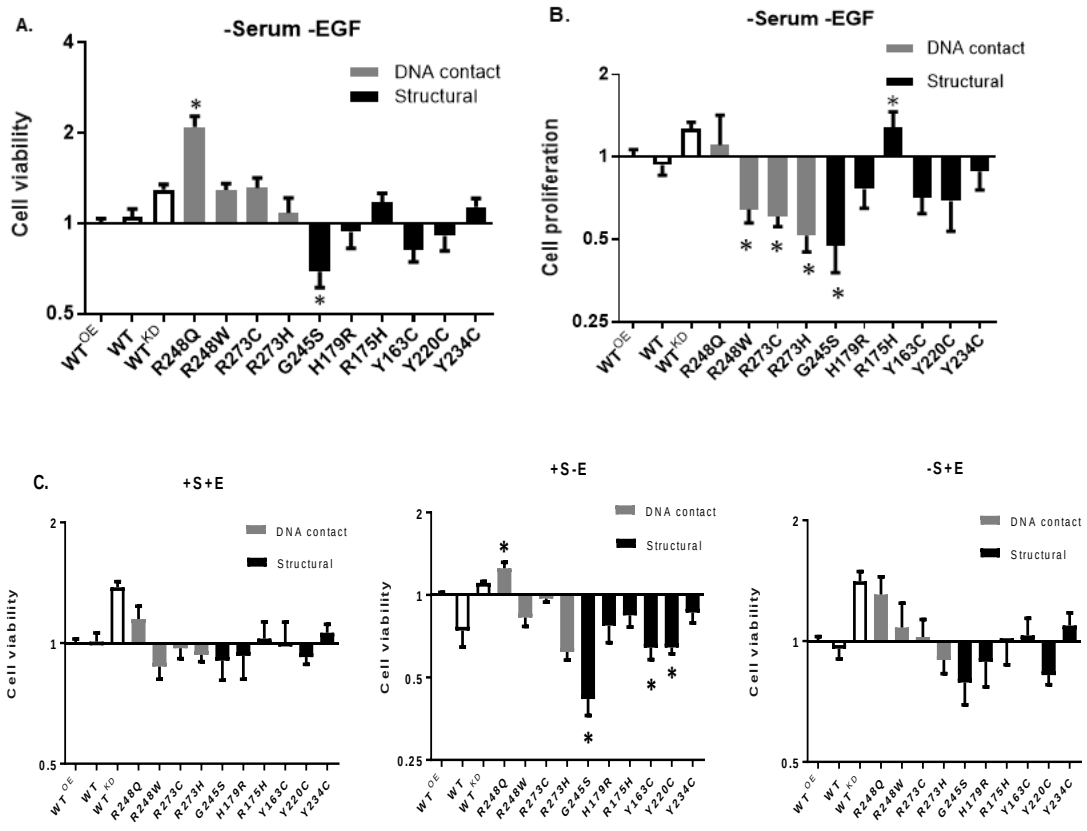


Figure 2.3. **Effect of p53 mutants on cell viability.** A. Viability of the MCF 10A p53 mutant cells in absence of growth factors- Serum and EGF (-S-E) measured by the CellTiter-glo assay. B. Cell counts measured by DAPI staining after 72 hrs growth in growth factor-deprived media. C. Cell viability measured by the CellTiter-glo assay after cells were either grown in the presence of both serum and EGF (+S+E) or in the absence of EGF (+S-E) or serum (-S+E). All the data have been analyzed from 3 separate experiments and normalized with respect to the phenotypic data of WT^{OE} cells, and the error bars show the SEM values.

Resistance to apoptosis

Apoptosis, the programmed cell death is a barrier to cancer development and is triggered by tumor suppressor *TP53*. Loss of function or mutation in *TP53* fails DNA damage sensing and hence, apoptosis induction (Aubrey 2018). Anthracyclins such as doxorubicin has been widely used as a primary chemotherapeutic drug and for adjuvant therapy for breast cancer to curb the proliferation and trigger apoptosis of the cancer cells. To test whether different missense p53 mutants make cells respond differently to the apoptotic stimuli, we evaluated the level of apoptosis in MCF 10A cells expressing p53 mutant proteins after 24 hrs of 1µM doxorubicin treatment by measuring the activities of caspase 3 and caspase 7 released during apoptosis, using a luminescent assay where luminescence was proportional to amount of caspase activity present.

Overall, MCF 10A p53 mutant cells were more resistant to apoptosis except for the mutants G245S and R273H compared to the $p53^{KD}$ cells. In contrast to the results for the cell viability assay in different growth conditions, the most apoptosis-resistant mutants were the structural mutants R175H, Y163C, Y220C, H179R and Y234C, while the other structural mutant G245S was significantly more sensitive compared to the WT^{OE} cell (Fig 2.4A, left graph). Among the DNA contact mutants, R248W was the most resistant to apoptosis followed by R248Q. However, overall, cell viability of mutant p53-expressing cells in the presence of 1µM doxorubicin (Fig 2.4A, right graph) was not significantly different from that of WT^{OE} cells. Similar results were found when we measured the fraction of apoptotic cells with annexin V staining by flow cytometry (Fig 2.4B). Annexin V stains the apoptotic cells by binding to phosphatidyl serine when it flips out on plasma membrane after cell death.

The tumors derived from $p53^{-/-}$ embryonic fibroblasts contained less apoptotic cells, and missense mutations in p53 were associated with resistance to multiple doses with gamma radiation or doxorubicin (Lowe 1994). In another study of 63 breast cancer patients receiving doxorubicin

monotherapy, 18 patients had mutation in the *TP53* gene, and mutations in the L2 and L3 loop regions in the DNA binding domain where the hotspot mutations are located, resulted in progressive disease during doxorubicin treatment. Specific p53 point mutation at codon 248 was implicated in primary resistance to doxorubicin and early relapse in breast cancer patients (Aas 1996). In agreement, our experimental studies showed that R248W was the DNA contact mutant resistant to apoptosis. In another study with p53 mutant-expressing H1299 lung adenocarcinoma cells, p53His175 and p53His179 were found to be resistant to etoposide-induced apoptosis while p53His273 along with p53Trp248 had a protective effect (Blandino 1999).

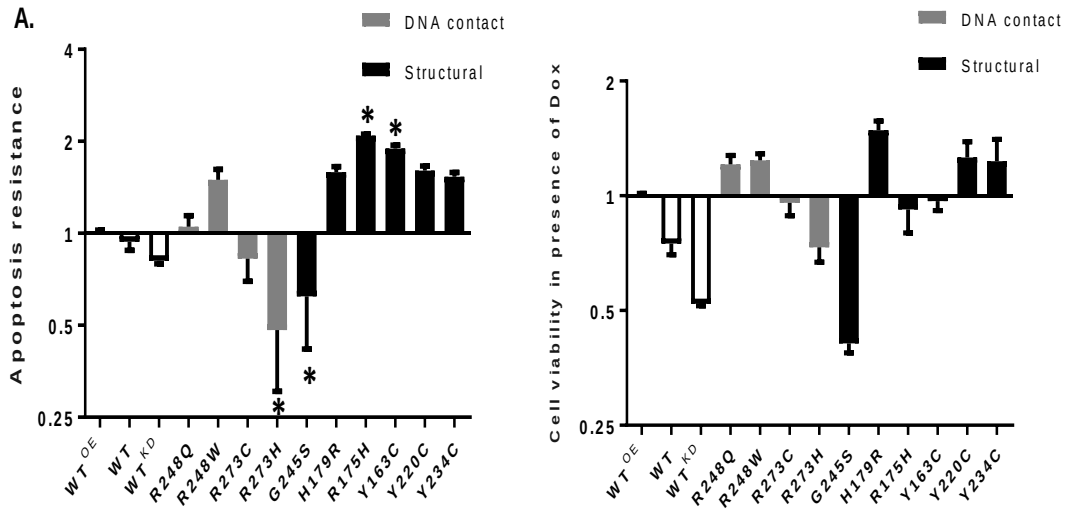


Figure 2.4. **Resistance to apoptosis of MCF 10A p53 mutants in presence of 1μM doxorubicin for 24 hrs.** A. Resistance to apoptosis measured by activity of caspase 3 and caspase 7 using the Apotox Glo kit (left graph). In a parallel cell viability assay in presence of 1uM doxorubicin show the structural mutants to be more viable along with mutants R248Q and R248W that correlated well with the apoptosis assay. R273H and G245S were not viable in doxorubicin. All the data have been analyzed from 3 separate experiments and normalized with respect to the phenotypic data of WT^{OE} cells, and the error bars show the SEM values.

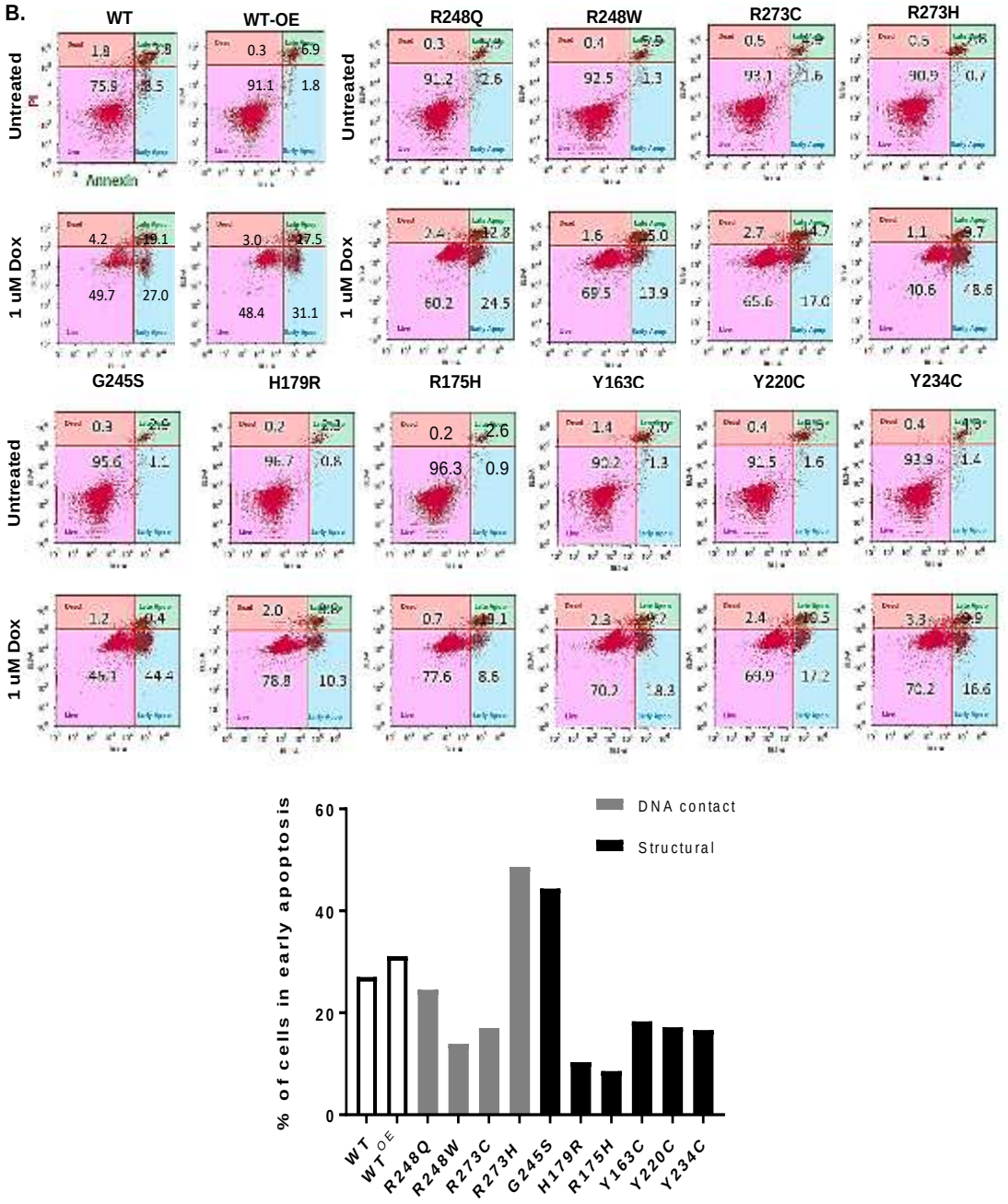


Figure 2.4. **Resistance to apoptosis of MCF 10A p53 mutants in presence of 1μM doxorubicin for 24 hrs.** B. Flow cytometry data for apoptosis induction in mutants after 24 hrs doxorubicin treatment and shows percentage of cells entering early apoptosis. These results were comparable to that found by measuring the caspase 3 and caspase 7 levels. All the data have been analyzed from 3 separate experiments and normalized with respect to the phenotypic data of WT^{OE} cells, and the error bars show the SEM values

Cell migration and invasion

One of the hallmarks of progression towards a metastatic phenotype is the ability of the malignant cells to migrate from the primary site, survive outside of their natural adhesion niche and colonize the secondary sites. This whole process involves several steps including migration of cells, local invasion in their immediate habitat, crossing the basement membrane, entering the bloodstream and finally invading and proliferating at a new secondary site (Powell 2014). Wildtype p53 protein negatively modulates cell spreading, polarization, and filopodia formation thereby inhibiting initiation of migration and invasion. On the contrary, mutant p53 protein enhances random cell migration and invasion. During the process, the adherent epithelial cells undergo transition from polarized epithelial cells to motile, long and spindle-like mesenchymal cells, which is known as the 'epithelial-to-mesenchymal transition' (EMT) that allows the cells to cut through the basement membrane and squeeze through the extracellular matrix (ECM) (Kalluri 2009). Wildtype p53 protein suppresses EMT by repressing the transcription factors SNAIL, SLUG and TWIST which are the master regulators of EMT.

To determine the migratory potential of the MCF 10A p53 mutant cell lines, we used a well-established cell migration assay based on the transwell chamber also known as the Boyden chambers. These chambers have a membrane with pores with 8 µm diameter, allowing only the slender migratory cells to cross the membrane. After incubation for 22 hours, the migrated cells were stained with crystal violet and counted under the light microscope. The most migratory mutants were R248W, R273C and Y220C (Fig 2.5A). R175H and Y163C were moderately migratory and migrated more than the p53 null cells. The least migratory mutants included Y234C, G245S and R273H.

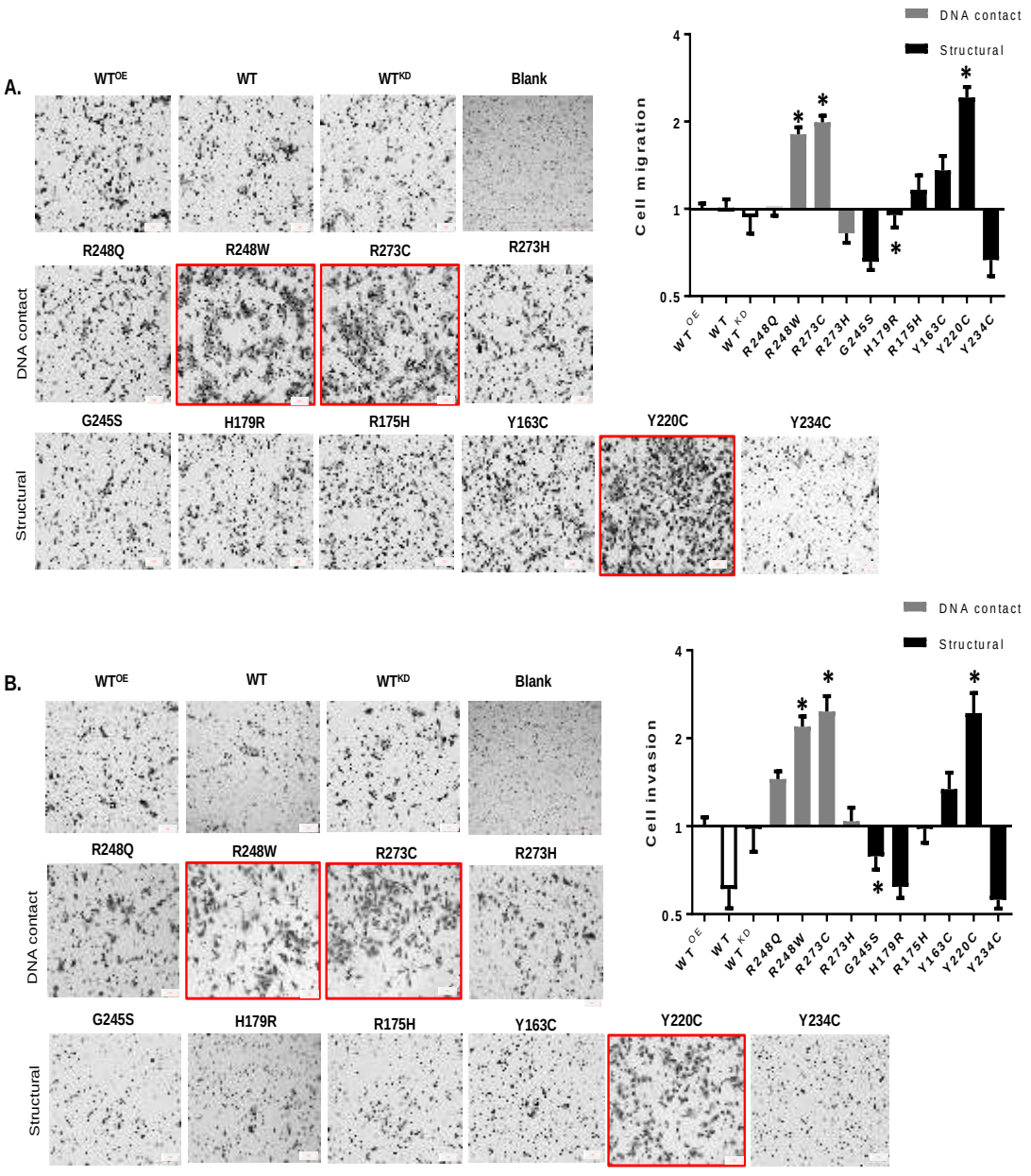


Figure 2.5. **Migration and invasion of MCF 10A p53 mutant cells.** A. Migratory potential of the MCF 10A p53 mutants. Mutants R248W, R273C and Y220C were the most migratory (highlighted in red). The least migratory mutants were G245S, Y234C and R273H. B. Invasion potential of the MCF 10A p53 mutants. R248W, R273C and Y220C (highlighted in red) were significantly more invasive than the WT^{KD} cells. The least invasive mutants were G245S, H179R and Y234C. The small black dots in blank well depict the pores in the membrane. All the data have been analyzed from 3 separate experiments and normalized with respect to the phenotypic data of WT^{OE} cells, and the error bars show the SEM values.

To measure cell invasion, the Boyden chamber membrane coated with Matrigel (reconstituted basement membrane), were used which needs to be proteolytically digested to allow invasion and thus, forms a perfect barrier for the non-invasive cells. The results provided a clear distinction between the non-invasive mutants G245S, H179R and Y234C and more invasive R248W, R273C and Y220C mutants. (Fig 2.5B), while R248Q and Y163C were moderately invasive. As the invasion process requires both cell migration and basement membrane breakdown, we compared the invasiveness and the migratory phenotype. The most migratory mutants R248W, R273C and Y220C were also invasive and crossed the Matrigel barrier. Similarly, the least migratory mutants Y234C and G245S proved to be non-invasive, while the mutant Y163C was moderately migratory and invasive. It is interesting to note that the moderately migratory mutants H179R and R175H failed to invade the Matrigel barrier, further demonstrating that not all cells that can migrate have invading potential, but cell migration is a pre-requisite for the invasion process. Lastly, it should be noted that even the most invasive R273C cells were far less invasive than the metastatic cells such as MDA MB-231 with additional mutations to *TP53* mutation, implying that p53 mutation by itself is not sufficient to induce the full invasive capacity.

To assess the EMT process during cell migration, we then performed the immunofluorescence staining for β-catenin (a marker for epithelial cells) and Vimentin (a marker for mesenchymal cells). We observed that the migratory cells, R273C and Y220C, had typical mesenchymal characteristics, including disorganized β-catenin (*i.e*., more in the cytoplasm than at the cell adherens junctions) and vimentin (*i.e.,* stretched across the cells) staining (Fig 2.5C). To further contrast the invasiveness of cells, two of the migratory mutant R175H (moderately migratory but not invasive) and R248W (highly migratory and more invasive) cells were seeded on a type I collagen bed with a Matrigel overlay to create a 3D matrix. The cells were then allowed to invade the Matrigel bed and imaged at different focal planes along the Z-axis. The R248W cells were detected up to 600um away from the base of the well, while very few R175H mutant cells could invade the Matrigel overlay (Fig 2.5D) restating that a migratory phenotype may not translate into an invasive phenotype.

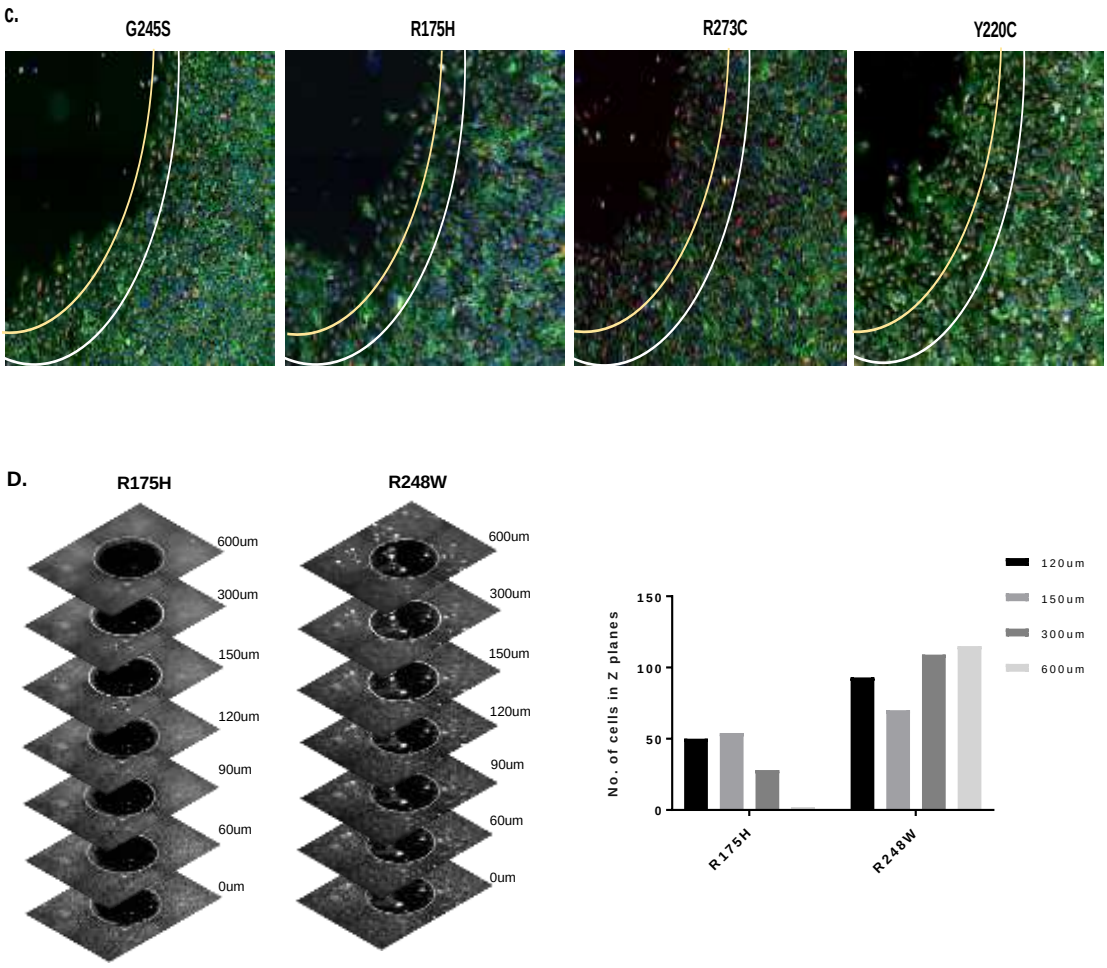


Figure 2.5. **Migration and invasion of MCF 10A p53 mutant cells**. C. EMT markers- β-catenin and vimentin staining in the least (G245S), moderately (R175H) and most (R273C and Y220C) migratory p53 mutants. Cells were seeded in wells with stopper. The white line shows the boundary of the stopper and the yellow line the extent of migration of the least migratory mutant G245S. β-catenin was observed to be delocalized to the cytoplasm from the cell junctions in the migratory cells. The migrating R273C and Y220C mutant cells looked very spread like a 'fan' which is typical for a migratory phenotype. Vimentin staining was also very prominent in the migratory cells (R273C). D. Invasion potential of R175H (less invasive) and R248W (highly invasive) mutants in 3D Matrigel matrix. R248W mutant cells were very invasive and invaded till 600um along the Z-axis. Very few R175H mutant cells could invade only till 150um along the Z-direction.

In previous reports, the mutant R273H has been shown to cooperate with p21 to promote Slug protein degradation and decreasing cell invasion (Kim 2014). *TP53* mutation, low MDM2 and high Slug expression correlates with poor survival and metastasis in NSCLC patients and expressing the p53-R248W mutant in the p53-null H1299 lung carcinoma cells decreased MDM2

protein levels and stabilized Slug to promote cell invasiveness (Wang 2009). *TP53* mutants are known to downregulate E-cadherin expression with simultaneous upregulation of transcription factors Slug and Zeb-1, which are known to repress E-cadherin expression. Loss of E-cadherin expression is accompanied by expression of the mesenchymal marker Vimentin and delocalization of β-catenin to cell cytoplasm, the characteristics of EMT (Roger 2010).

Moreover, we showed that different missense mutations at the same position (*e.g.*, R248W/Q and R273C/H) could affect cell migration and invasion differently. Similarly, a study with lung and breast cancer cell lines stably expressing R273H, R273C and R273G p53 mutant proteins demonstrated different gene expression profiles resulting in different phenotypes, where the R273H- and R273C-expressing cells were more migratory than the R273G-expressing cells (Li 2014). Overall, these findings demonstrated that a single amino acid change can have strikingly different effects on cellular phenotypes and hence likely the clinical outcomes.

Resistance to anoikis

When adherent cells are detached from extracellular matrix (ECM) attachment, an apoptotic process known as anoikis is initiated, and resistance to anoikis is a characteristic of transformed cells to enhance tumor metastasis (Gilmore 2005). Upon detachment from three-dimensional matrix, a p53 dependent apoptotic program is induced in fibroblast and endothelial cells, and loss of functional p53 protein enables cell survival without any anchorage dependency. p53 controls survival signal from ECM and FAK in anchorage dependent cells, and expression of dominant-negative forms of mutant p53 protein resulted in anoikis resistance (Ilic 1998).

To check the effect of different p53 missense mutant proteins on anoikis, the MCF 10A p53 mutant cells were grown in low-binding 96-well plates for 7 days to form the cell aggregates, mimicking the conditions *in-vivo* when cells enter the bloodstream after detachment from the basement membrane. These cell aggregates were then stained with EtBr and Vybrant violet to

stain dead (*i.e.*, EtBr-permeable) and live cells, respectively. When the death index (dead cells/ live cells) was calculated, the mutants G245S and Y234C were most sensitive to anoikis *(i.e.*, with higher death index) and underwent cell death in absence of attachment to any matrix (Fig 2.6). The rest of the p53 mutants were more resistant to anoikis than the WT^{OE} cells, whereas MCF 10A cells with endogenous wildtype p53 (WT) and with p53 knock-down (WT^{KD}) were sensitive to anoikis compared to the MCF 10A cells over-expressing wildtype p53 (WT^{OE}). The most resistant p53 mutants were R248W, R273C, R175H, H179R followed by Y220C where the spherical aggregates comprised mostly of live cells with a low death index in complete absence of any attachment to 3D matrix. The other mutants R248Q, R273H and Y163C also demonstrated anoikis resistance (more live cells than dead) but to a lesser extent than the most resistant mutants.

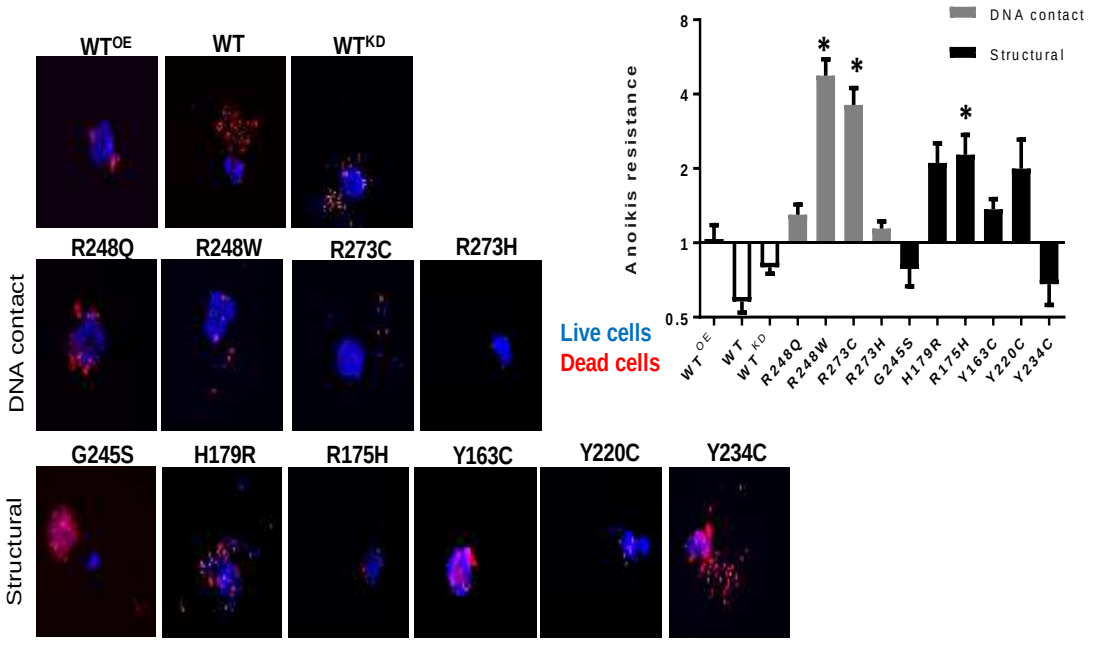


Figure 2.6. **Anoikis resistance of the p53 mutant MCF 10A cell lines.** Cell aggregates were stained with Vybrant violet dye (live cells, blue) and EtBr (dead cells, red), and the death index (dead/ live cells) was calculated. All the data have been analyzed from 3 separate experiments and normalized with respect to the phenotypic data of MCF10A p53 WT^{OE}. The error bars show the SEM values.

Supporting our results, breast cancer cells MDA-MB-468 harboring R273H mutation in *TP53* was reported to suppress anoikis (Tan 2015). EMT is known to confer anoikis resistance and increases the likelihood of survival of the invasive cells that breach the basement membrane and enter the blood flow (Cao 2016), and we observed that the most migratory and invasive R248W, R273C and Y220C cells were also resistant to anoikis. However, the R273H cells were not invasive but resistant to anoikis, indicating an alternative mechanism of resisting anoikis in the absence of EMT.

The 3-dimensional mammosphere morphology

The MCF 10A mammary epithelial cells form spherical, growth arrested, polarized structures in three dimensional Matrigel matrix (ECM), the mammospheres, that resemble the glandular epithelial architecture of mammary acini (Debnath 2003). The mammosphere has a clear lumen (*i.e*., the absence of cells at the center of the spheroids) due to anoikis or apoptotic death of the centrally located cells devoid of attachment to the substratum. The resultant spheroid looks like a hollow ball, with a single layer of cells only at the periphery. The over-expression of wildtype p53 does not affect this phenotype but cells without p53 (p53 null) and cells with the hotspot *TP53* mutations are known to develop non-polarized, disorganized cell clusters without a hollow lumen, which is either filled or partially cleared, indicating transformation of cells. Malignant tumor cells such as T4-2 form highly disorganized and continuously proliferating colonies in 3D Matrigel matrix (Kenny 2003).

For high-throughput assessment of mammosphere formation in a 96-well format, we developed the ‘on top’ method where cells were seeded on top of a Matrigel layer rather than being embedded. This method allowed reduced time for acini formation (~7 to 9 days) and clearance of lumen as well as similar focal planes for automated confocal Z-stack imaging compared to the embedded method. An image analysis pipeline was also developed using the CellProfiler software to quantify i) the spheroid sizes ii) number of cells in the lumen (counting nuclei in the

lumen) or hollowness of the mammospheres iii) expression of cell polarity marker like laminin (for apico-basal polarity) by relative fluorescent antibody intensities.

The size of the spheroids was determined by measuring the diameter of the spheroid cross-sections and applying the formula for the area of a circle after a pixel-to-µm conversion.
When we tested the cells expressing different missense p53 mutant proteins for mammosphere fomation, the MCF 10A p53 WT^{OE} cells formed the largest spheroids with a mean area of 785 μm^2, whereas the R273H cells formed the smallest dense spheroids with a mean area of 65 μm^2 (Fig 2.7A). The MCF 10A cells with endogenous wildtype p53 and $p53^{KD}$ control had comparable mean areas of 450 μm^2 and 470 μm^2, respectively. All the cells expressing DNA contact mutants had mean spheroid areas significantly smaller than the $p53^{OE}$ and $p53^{WT}$ spheroids, with mean areas ranging from 65 μm^2 to 266 μm^2. The cells with structural mutants had spheroid areas comparable to that of $p53^{WT}$ control, with an exception of the G245S cells that had the second smallest spheroid (mean area157 μm^2). The mean spheroid sizes for the p53 structural mutants were twice as large than the DNA contact mutants. It was also interesting to note that the spheroids derived from cells expressing the two most invasive mutants show a significant size difference, where the R248W and Y220C spheroids had areas of 266 μm^2 and 475μm^2, respectively. While both the mutants R248W and Y220C behave similarly in the invasion chambers penetrating a thin Matrigel layer, they develop different morphologies in the 3D Matrigel alluding to other significant changes between them. The p53 mutants R248Q, R175H and Y220C formed deformed spheroids of various shapes.

The MCF 10A p53 WT^{OE} cells formed the mammary acini with cleared lumen like the $p53^{WT}$. All p53 mutants had cells at the center of the mammospheres. The mutants with most prominent nuclear staining indicating a filled lumen were the DNA contact mutants (Fig 2.7B). The DNA contact mutant R273H and the structural mutant G245S showed a substantial nuclear stain intensity due to the tiny and dense spheroids. The DNA contact mutants had filled lumens and structural mutants had partially filled lumens compared to the $p53^{WT}$ and p53 WT^{OE}

mammospheres. p53^{KD} did prove to elicit the abnormal spheroid morphology with the nuclei staining intensity in the inner bins like p53 mutants R248W and R273C.

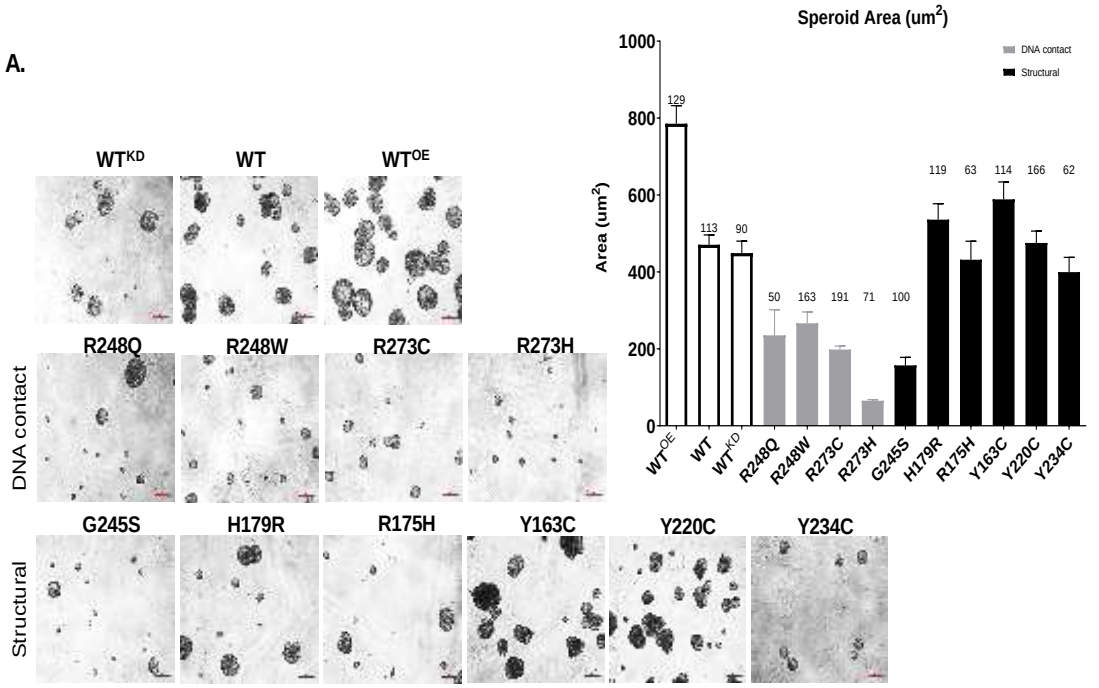


Figure 2.7. **Morphology of the MCF 10A p53 mutant mammospheres.** A. The spheroid sizes were measured on day 9 on the on-top culture in 3D Matrigel bed. The WT^{OE} formed the largest spheroids. Some of the mutants had a mixed population of small and big spheroids (R248Q, G245S, H179R and R175H) but on average G245S formed smaller spheroids. All the data have been analyzed from 3 separate experiments and normalized with respect to the phenotypic data of MCF10A p53 WT^{OE}. The error bars show the SEM values.

Another important morphological characteristic of normal MCF 10A mammary spheroids is the polarization of the cells in the single peripheral layer, where the apical side orients towards the lumen (center of the mammosphere) and basal side away from the central lumen marked by the protein markers. We measured the localization of laminin (Fig 2.7C), a protein that is present on the basal side (outer side) of the spheroids lining the circumference of the hollow sphere. To quantify the formation of properly structured spheroids, we developed a method of estimating the 'hollowness' of mammospheres using the CellProfiler by taking the equatorial cross-section pictures and measuring the staining intensity across 20 concentric rings with increasing radius,

called bins, expanding outward (Fig 2.7C). In this setting, the bins 1 and 20 represent the innermost and outermost rings, respectively, and higher signal intensity of laminin in the inner bins indicates reversed polarity of cells. The intensity value data was normalized to the total number of pixels for each spheroid since the larger the bin, the more pixels present.

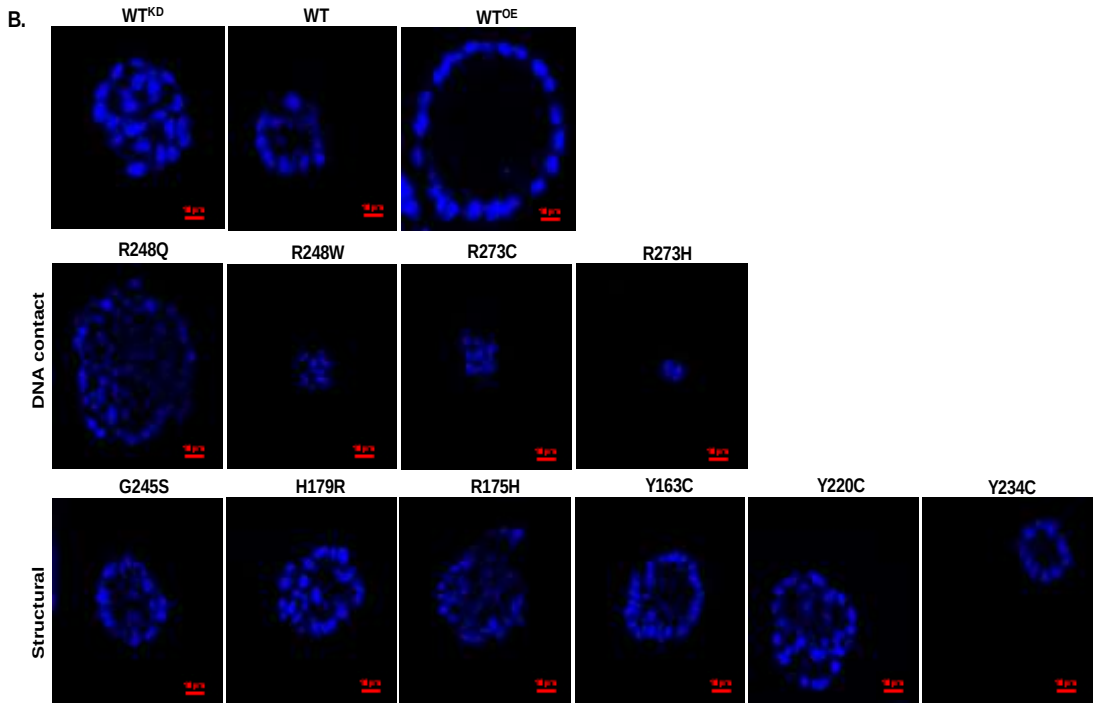


Figure 2.7. **Morphology of the MCF 10A p53 mutant mammospheres**. B. DAPI staining showed lumen clearance of mammospheres. None of the mutants could form the normal hollowsphere, and DNA contact mutants formed quite dense spheres compared to the structural mutants.

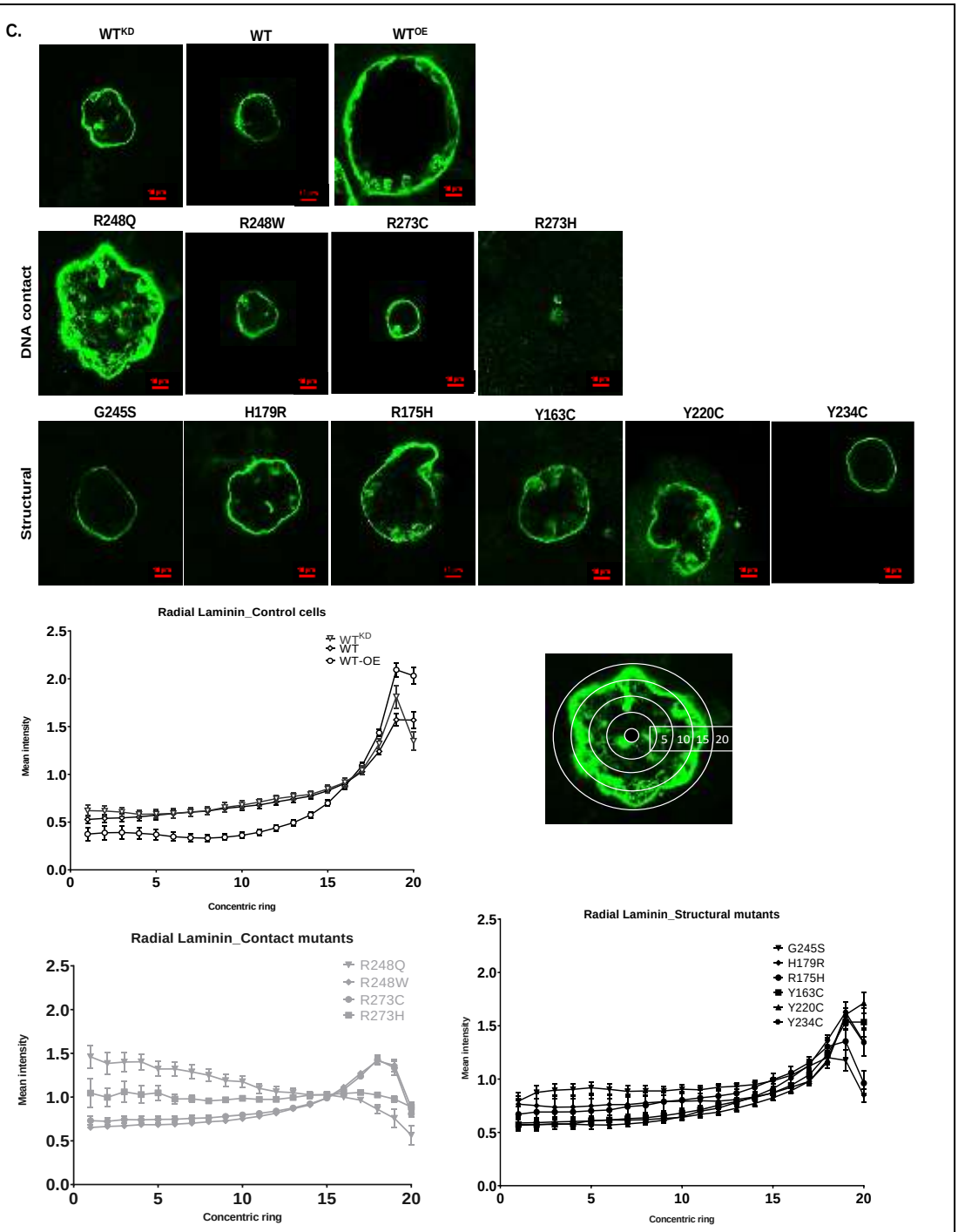


Figure 2.7. **Morphology of the MCF 10A p53 mutant mammospheres**. C. Laminin staining intensity profiles across the bins. The error bars represent SEM from independent measurements of an average of spheroids.

The $p53^{OE}$ and $p53^{WT}$ cells demonstrated laminin staining on the outer side of the spheroids indicating normal cell polarity (Fig 2.7C). The $p53^{OE}$ cells had laminin staining not detectable until

bin 13 (out of 20) that was close to the periphery, which then steadily increased until bin 20. The $p53^{WT}$ cells had no central laminin intensity inside the spheroids, while laminin staining inside the spheroid was detected for $p53^{KD}$ cells suggestive of disrupted polarity of cells. The missense mutant with a strong laminin intensity towards the center of the spheroid was R248Q (Fig 2.7C). The laminin intensity of R273H and G245S cell was constant from the spheroid center to the periphery because of the small and dense spheroids (the bin size is proportional to the size of the spheroid as a whole). Among the mutant p53 expressing cells, only the R248Q cells showed a consistent decrease in laminin staining intensity from the center towards the periphery indicating reversed polarity of cells. The two invasive mutants, R248W and Y220C, also led to a decreased laminin staining at the periphery. In addition, we stained the mammospheres for β-catenin to assess the status of EMT, and observed that the cells expressing Y220C and R248Q showed disrupted β-catenin staining (Fig 2.7D), indicating altered cell polarity via EMT.

Earlier work has shown that knock-down of endogenous wildtype p53 in MCF 10A cells form mammary acini with partial clearance of the lumen and altered staining of β-catenin and laminin V (Zhang 2011). In agreement with our results, ectopic expression of p53 mutants G245S, R248W, R175H and R273H formed disrupted acini with filled lumen along with increased expression of mesenchymal markers like Snail, Slug and Twist and decreased levels of epithelial markers and staining patterns of E-cadherin, β-catenin and ZO-1. Thus, collectively with previous reports, our data suggest that a subset of missense mutant p53 caused change in cell polarity through EMT.

Overview of phenotypic characteristics of MCF 10A cells expressing mutant p53 protein

In summary, we have developed a series of high-throughput quantitative assays to assess the cancer hallmark phenotypes of normal mammary epithelial MCF 10A cells over-expressing the ten most common p53 missense mutant proteins found in breast cancer. To normalize the effect of protein overexpression in the context of identical genetic background, the phenotype scores for each mutant was calculated with respect to the phenotype of the MCF 10A cells overexpressing

the wildtype p53. The carefully designed and performed experiments provided a spectrum of mutant-specific phenotypic changes occurring across the ten p53 missense mutant protein expressing mammary epithelial cells and established that the different mutant proteins are functionally unequal (Fig 2.8). Each of them produces a 'neo-morphic' phenotype different from wildtype p53 and $p53^{KD}$ cells, which may reflect the heterogeneity of phenotypes observed in patient tumor samples.

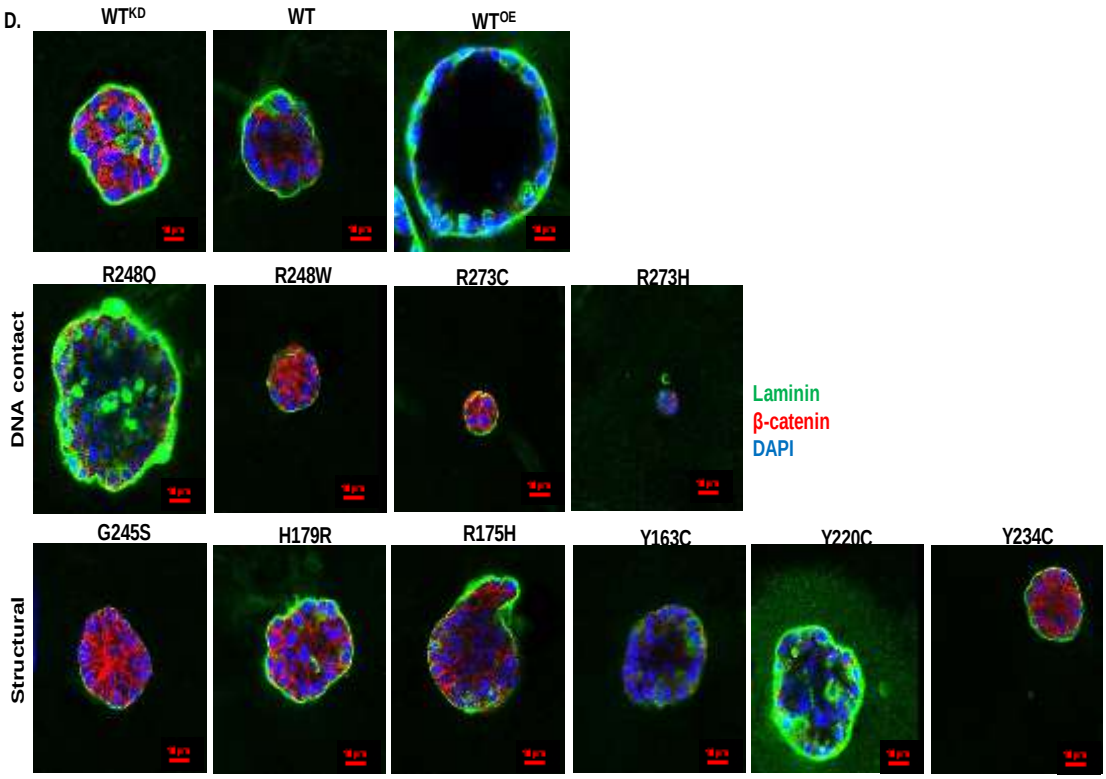


Figure 2.7. **Morphology of the MCF 10A p53 mutant mammospheres**. D. Matrigel Confocal images of equatorial cross-section of mammospheres stained with laminin, β-catenin and DAPI.

Based on the assessed cancer hallmark phenotypes, the highly aggressive p53 mutants were the two DNA contact mutants R248W and R273C and a structural mutant Y220C (Fig 2.8). These

aggressive mutants were viable under growth factor deprivation (absence of EGF and serum) and were resistant to apoptosis under doxorubicin treatment. Apoptotic resistance of the R273C cells was comparable to wildtype $p53^{KD}$ cells whereas the R248W and Y220C cells were more resistant to apoptosis than R273C and $p53^{KD}$ cells. Cells expressing R248W, R273C and Y220C mutants were the most migratory and invaded the Matrigel barrier establishing themselves as the most invasive mutants. These invasive mutants survived without attaching to a substratum demonstrating anoikis resistance that correlated with their resistance to apoptosis. In 3D Matrigel bed, cells expressing the anoikis-resistant DNA contact mutants R248W and R273C formed very small and dense mammospheres unlike the large hollow spheres with a single layer of polarized cells formed by mammary epithelial cells with wildtype p53. Cells expressing the Y220C structural mutant formed mammospheres of distorted shapes with partially filled lumen, and more laminin staining was observed inside the spheres than on the periphery, which indicated disruption of cell polarity that correlated with the migratory and invasive characteristics.

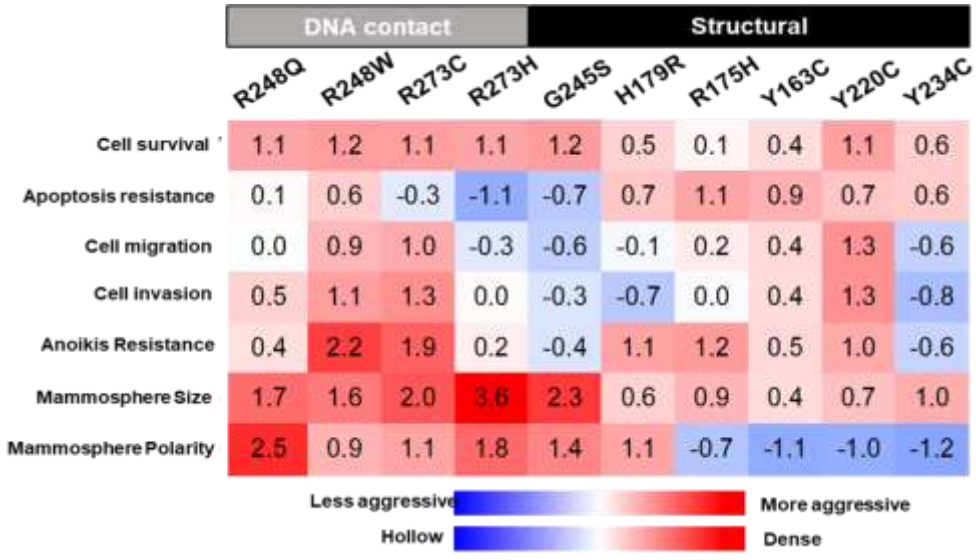


Figure 2.8. **Heat map of the phenotypic characteristics of the ten MCF 10A cells over-expressing p53 missense mutants based on the hallmarks of cancer.** The phenotype scores were calculated with respect to the results for MCF 10A wildtype p53 over-expressing cells.

An exceptional missense mutant was G245S with phenotype comparable to that of the MCF 10A cells overexpressing wildtype p53. The G245S cells had decreased viability in the absence of serum and EGF and was sensitive to doxorubicin treatment undergoing apoptotic cell death. The G245S cells were neither migratory nor invasive and did not show anoikis resistance.

Taken together, while the rest of the mutant expressing cells displayed intermediate cancer hallmark phenotypes, we can position G245S on one end of the spectrum as the least aggressive mutant and R248W, R273C and Y220C on the opposite end as the most aggressive mutants. Based on these distinct phenotypic patterns, we hypothesized that diverse phenotypes arose from neo-morphic transcriptional function of each mutant p53 protein, which then led to downstream changes in gene expression profile and cellular pathways. Therefore, to obtain genome wide molecular profiles, we next performed the RNA seq and ChIP seq, followed by comprehensive bioinformatics analyses.

RNA seq data and pathway analysis

To profile the baseline transcriptome, we performed RNA Seq on the MCF 10A cell lines expressing missense mutant p53 proteins as well as the cells overexpressing wildtype p53 ($p53^{OE}$). We extracted the total RNA from the cells, enriched for mRNA and sequenced on the Illumina HiSeq 2000 platform. Our goal was to understand the underlying molecular basis of the phenotypic heterogeneity of the p53 mutants.

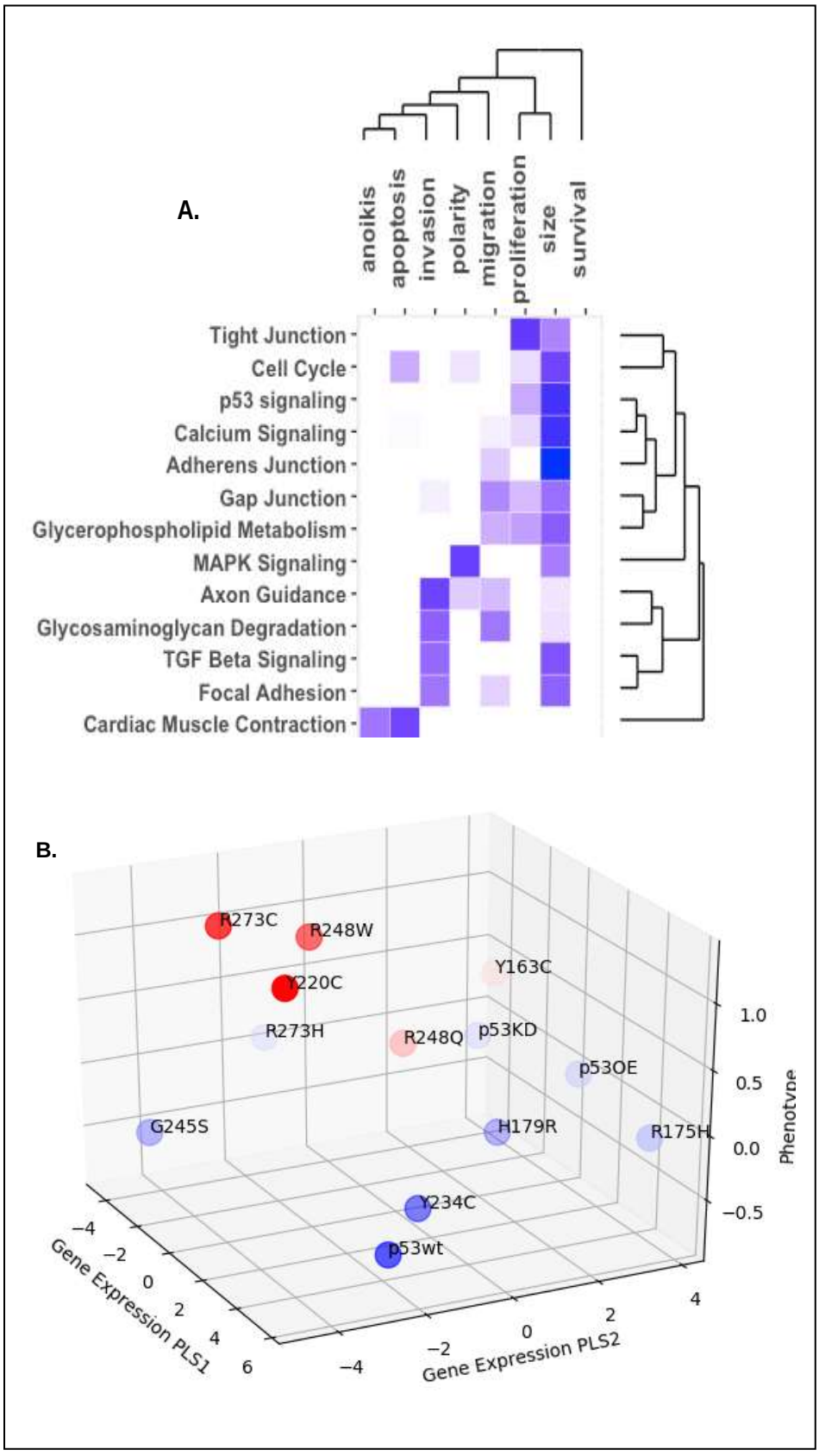
After aligning the reads using STAR (Dobin 2013), we identified an average of 24,252 genes and microRNAs with 20X coverage and normalized to TPM (Transcripts Per Kilobase Million) in all cell lines. Next, we carried out PCA (Principal Component Analysis) and pathway analysis to systematically analyze the genes differentially expressed among the cell lines (mutant p53 vs wildtype p53). For PCA, we selected the top 1,000 most variant genes across all 13 cell lines (10

p53 mutants and 3 controls) and performed dimension reduction to extract the fundamental structure of RNA expression level then mapped them onto a 2-D space (Fig 2.9D). Y234C and R273H were separated from the other cell lines, which indicates general differences in expression level. For pathway analysis, correlation between expression levels and phenotypic values across 13 cell lines (10 p53 mutants and 3 controls) were calculated for each gene and phenotype, and top 500 positively or negatively correlated genes were identified. By using the ontology and pathway terms in DAVID and Reactome databases, pathway enrichment tests were performed by the Fisher's exact. Because conventional pathway enrichment analysis are based on the counts of genes and don't take gene expression or continuous variables of phenotypes into account, we applied model-based approaches to identify the phenotype-associated pathways, in which expression levels of the genes were informative in explaining the phenotypic differences such as cell invasiveness (Fig 2.9). We specifically used a classification method (Pathifier) and a regression method (Partial Least Squares regression or PLS) for pathway analysis.

Pathifier (Drier 2013) performs non-linear regression to find a principal curve spanning the data points (*i.e*, cell-lines in our data) after dimensional reduction on the expression levels of a given set of pathway genes by the principal component analysis (PCA). Based on the relative position on the curve, a dysregulation score is assigned to each sample, and the scores are compared between pre-defined groups (*e.g*, cells with higher and lower invasiveness) by statistical tests such as a non-parametric Mann-Whitney-Wilcoxon test to determine whether the expression levels of the pathway genes are informative with statistical significance in classifying two groups of cells. When we tested the pathways in KEGG for weakly invasive cells (G245S, Y234C, H179R, R175H, p53-wt, p53-KD) vs. more invasive cells (R273H, Y220C, R273C, R248W, p53OE, Y163C, R248Q), metabolism-related pathways, cell adhesion related pathways, notch signaling, and axon guidance were identified as the most dysregulated pathways across the least to the most invasive mutant p53 expressing cells (Fig 2.9A).

To avoid any bias in the data, we applied PLS, a multivariate statistical regression method, which builds robust linear models on continuous response variables and particularly suitable for high dimensional data with a small sample size. With PLS, we can build models on individual pathways to test whether the expression levels of pathway genes (predictor variables or training vectors) can explain the continuous trend of phenotypes (response variable or targeting vectors) by measuring the coefficient of correlation of the models. From the analysis, we identified that the genes sets in pathways like Hippo signaling pathway (correlation = 0.71) , ECM-receptor interaction (correlation = 0.56), and focal adhesion (correlation = 0.52) in KEGG; ERBB4 signaling (correlation = 0.76), and Ca^{2+}/K^{+} channels (correlation = 0.67) related terms in Reactome can explain the trend of cell invasion observed across the ten p53 mutant cell lines. When the regression results were projected onto the PLS components, for the focal adhesion pathways, the most invasive p53 mutants (R273C, Y220C and R248W) as determined from the Matrigel-coated transwell invasion assay clustered together while the least invasive mutants H179R and Y234C clustered on the other end (Fig 2.9B). To highlight the genes that contribute most to the phenotypic differences, we overlaid the pathway diagram with the correlation between the expression levels of each gene and the phenotypic scores (Fig 2.9C). In addition, to show the potential direct transcriptional targets of p53 proteins, we also highlighted the known p53 target genes as well as the genes that we identified from our ChIP Seq data for p53 WT and mutant p53 proteins.

To select the genes for further validation, we calculated the invasion correlation scores by Pearson correlation between RNA expression level and invasion level among the 13 cell lines. We then ranked the scores to get a list of top genes who exhibit high RNA expression level in invasive cell lines while low in non-invasive cell lines (Appendix B). The top list of genes includes STAP2, CYP4B1 and PARP10 (high in invasive p53 mutants) and PLPPR2, SREBF1 and PDK2 (high in non-invasive p53 mutants).


A.
anoikis
apoptosis
invasion
polarity
migration
proliferation
size
survival
Tight Junction
Cell Cycle
p53 signaling
Calcium Signaling
Adherens Junction
Gap Junction
Glycerophospholipid Metabolism
MAPK Signaling
Axon Guidance
Glycosaminoglycan Degradation
TGF Beta Signaling
Focal Adhesion
Cardiac Muscle Contraction
B.
R273C
R248W
Y163C
Y220C
R273H
R248Q
p53KD
p53OE
G245S
H179R
R175H
Y234C
p53wt
Gene Expression PLS1
Gene Expression PLS2
Phenotype

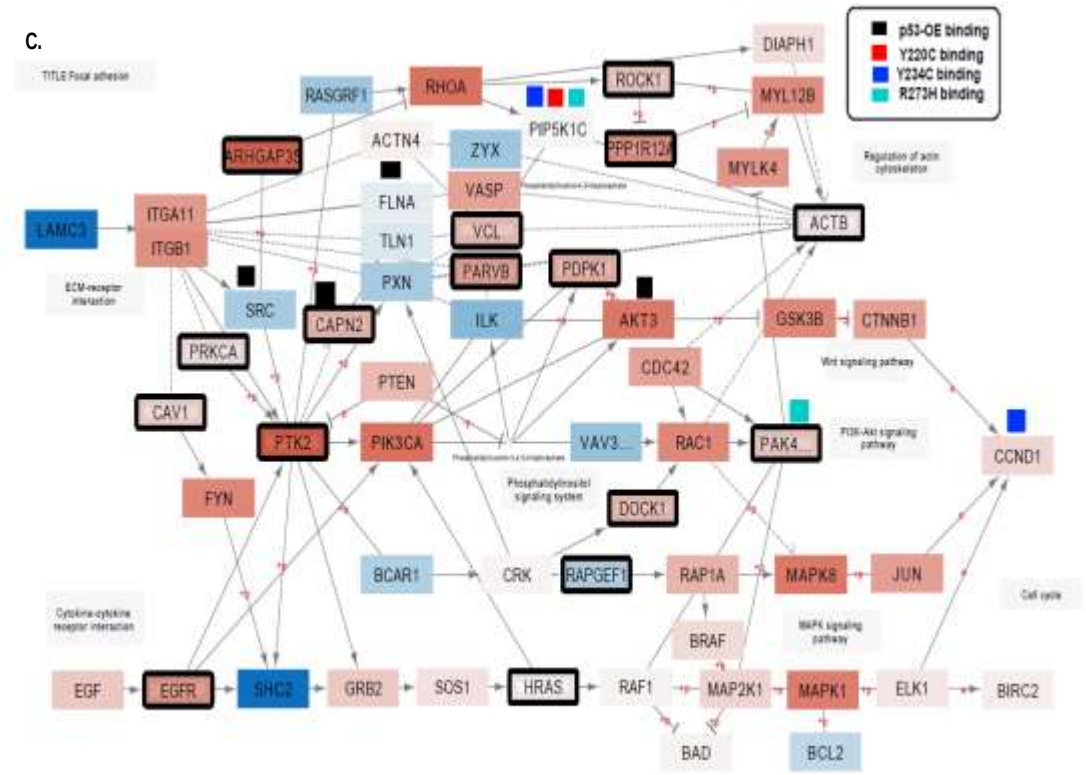


Figure 2.9. **RNA seq and pathway analysis of p53 mutants.** A. Pathway heatmap. Top ranked pathways in PLS analysis. B. PLS 3D plot. RNA expression scores of 24,252 genes were reduced to 2 dimensions (x and y axis). Cell lines of high/low (red/blue) invasiveness are separately clustered in the Focal Adhesion pathway. C. Focal adhesion pathway visualization. Red and blue in gene boxes indicate the high/low RNA expression level based on cell invasion level. Black borders imply known p53 targets. ChIP binding in MCF 10A p53OE, Y220C, Y234C, R273H are highlighted. D. PCA 2D plot. 1,000 genes were selected ranked by coefficient of variation across 13 cell lines.

ChIP seq data

The p53 protein is a sequence specific transcription factor, and to understand the DNA binding patterns of different missense mutant p53 proteins, we performed ChIP Seq on a set of two most invasive p53 mutants R273C and Y220C as well as two least invasive p53 mutants R273H and Y234C cells. To match the experimental conditions for RNA Seq, we used untreated cells (no treatment to induce high expression of p53 protein) and immunoprecipitated the C-term V5-tagged mutant p53 proteins expressed from the genome integrated plenti4 V5-DEST cassettes using antibody against the V5 tag. As a positive control we included the MCF 10A $p53^{OE}$ cells, overexpressing wildtype p53 with a V5 tag, while the parental MCF 10A cells expressing endogenous wildtype p53 (without any V5 tag) was used as the negative control for detecting the ChIP Seq peaks. Unlike the conventional p53 ChIP Seq typically performed after treatment (with Doxorubicin) of cells that elevate the cellular p53 protein level, we used untreated cells. Therefore, when we examined the enrichment of known DNA binding motifs, we did not observe a strong over representation of the p53 binding motif in the identified peaks in the $p53^{OE}$ cells, in contrast to the studies using Doxorubicin treatment where > 30% peaks contained p53 binding motif (Tonelli 2017). In our dataset, 7.91% of binding sites at promoter region were identified as p53 binding motifs which is similar to a previous study performed with the WT p53 (Idogawa, 2014), indicating a limited basal level of DNA binding of WT p53 in the absence of any stress.

Different DNA binding capacity and preference of p53 missense mutant proteins

Across the WT and 10 mutant p53 expressing cell lines, the p53 binding sites were mainly identified at the intergenic region (33% - 80%), and 5% - 23% of peaks were located within the promoter regions (+100/-2,500 bps of transcription start sites). Although no universal binding motifs were identified for both WT and mutant p53 proteins within the promoter regions by the *de*

novo motif discovery tool Homer at the significance level of $P < 0.001$, several distinct binding motifs were observed for the different p53 mutants, implying the 'neo-morphic' functional variation across different p53 mutations. Upon examining the enrichment level of known p53 binding motifs at the promoter region, it was evident that the mutant forms of p53 (especially the invasive R273C) showed reduced binding to the canonical p53 motifs, while the degree of reduced binding did not correlate to the invasiveness of cells (Fig 2.10A). Well-known targets of wildtype p53 such as CDKN1A were pulled down only by WT $p53^{OE}$ but not by the mutant p53. When we compared the binding motifs for all p53 mutants, R273H showed most similar binding preference as the WT $p53^{OE}$, while R273C bound to a very small number of distinct DNA motifs from WT and other p53 mutants (Fig 2.10B), implying general loss-of-function of DNA binding.

In addition, compared to the invasive mutants of Y220C and R273C, the less invasive mutants R273H and Y234C presented more similar motif binding pattern as wildtype p53 protein for both known (Fig 2.10A) and *de novo* motifs (Fig 2.10B). Overall, these indicate that the missense mutant p53 proteins, especially the more invasive ones, had decreased affinity for canonical p53 binding motif and altered binding preference (neo-morphic function).

Further, in agreement with a previous study that suggested an alternative mode of transcriptional regulation of the p53-R273H mutant without direct DNA binding via interacting with a transcription factor ETS (Vaughan 2014), the binding motif for ETS was over-represented at promoter regions targeted by R273H mutant. Moreover, the most prevalent ETS binding motif, CCGGAAG, was found in 22% of promoter peaks with p-value $< 10^{-7}$ in R273H, and this binding motif was over-represented for all mutant p53 proteins when compared to $p53^{OE}$.

Taken together, these suggest that mutant p53 proteins may exert neo-morphic transcriptional functions both by changing DNA binding preference and via interaction with other transcriptional factors.

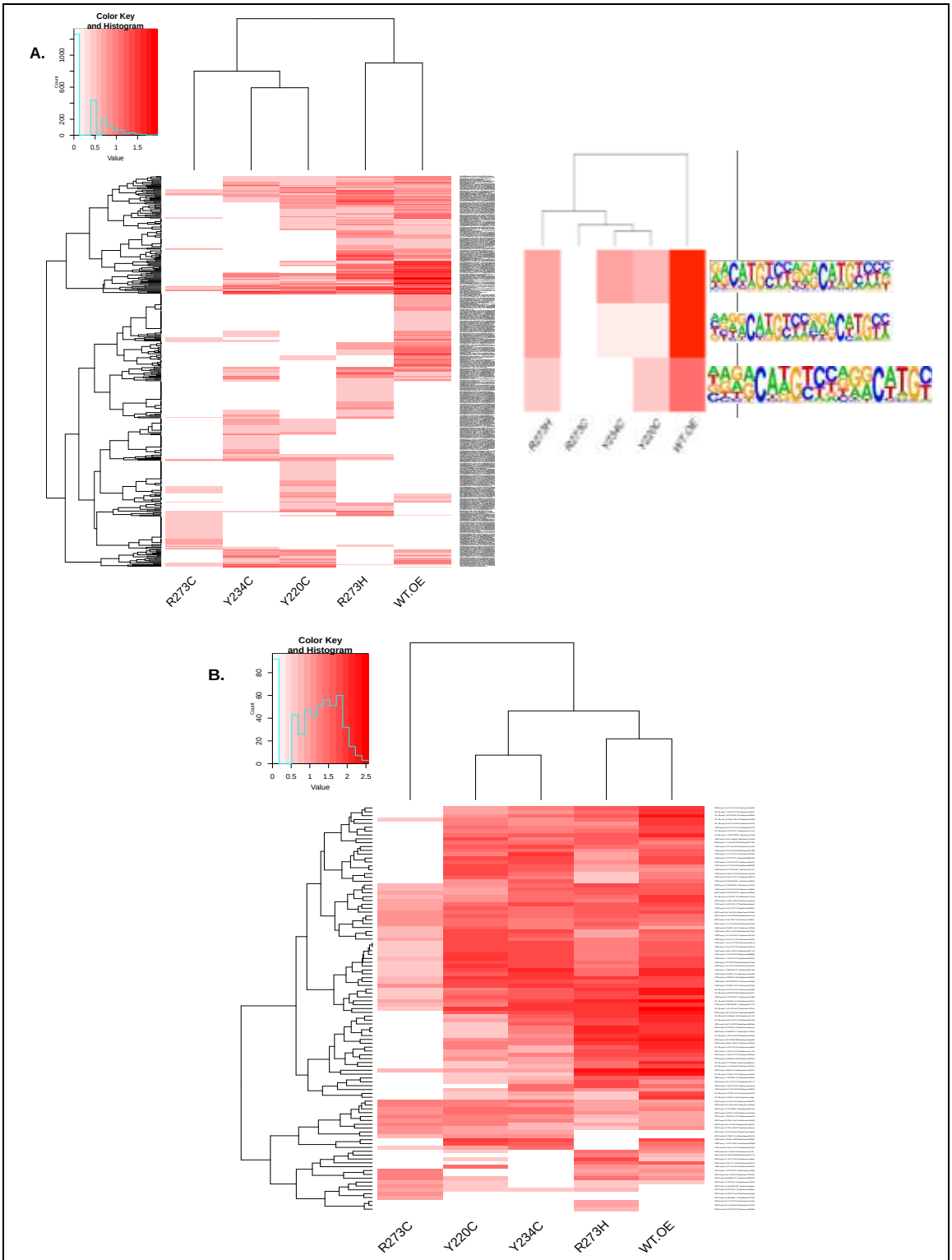


Figure 2.10. **Results from ChIP seq.** A. Figure of known p53 motif heatmap. Distinct known motif binding was observed in different p53 mutant cell lines. B. De-novo motif heatmap. Distinct de-novo motif binding were observed in different p53 mutant cell lines. The p53 mutant cell lines lose the function of canonical sequence specific DNA binding.

Inferring transcriptional activities of different p53 mutants by integrated analysis of ChIP Seq and RNA Seq

The p53 can function as both a direct transcriptional activator and an indirect transcriptional repressor. We thus, examined how the promoter binding affected the expression of target genes by comparing the expression level in each cell to the average levels in other cell lines. In general, weakly invasive mutants targeted considerably more genes and induced broader and more significant expressional changes (Fig 2.11). Further, when we examined the expression changes in each cell line, a large fraction (66%) of genes differentially expressed by R273H (but not Y234C) were direct binding targets, indicating more active transcriptional function of R273H upon binding to its targets. In contrast, much less influence on gene expression of strongly invasive mutants implied alternative modes of driving aggressiveness (*e.g.,* via protein interaction with ETS). Overall, our integrated analysis revealed both loss and gain of DNA binding function of different mutant p53 proteins, which in part explains the phenotypic heterogeneity.

Though the overall trend of expression direction for each p53 mutant was not statistically significant in GSEA test, expressions of many target genes were differentially regulated. For example, we identified genes such as *TFAP2A*, whose promoter region was occupied by Y234C and displayed a 2-fold increase in RNA expression. Overexpression of this gene decreases cell migration and invasion (Huang 2016), which corresponded with the phenotype of Y234C as the least invasive p53 mutant among the ten chosen p53 missense mutants. A negative correlation with RNA expression level and invasion/migration level was also observed. p53-Y234C protein bound to promoter of *CASZ1*, a zinc finger transcription factor that functions as a tumor suppressor, and its expression level dropped to 0 in Y234C cells. Interestingly, in the TCGA data, breast cancer patients with *CASZ1* mutation have a significantly shorter survival time ($P < 0.05$). We also identified genes such as *BCKDK*, whose promoter was occupied by mutants R273H and Y234C and displayed a 4-fold increase and 2-fold decrease in RNA expression respectively. BCKDK can enhance MAPK signaling pathway through direct MEK phosphorylation (Xue 2017)

and promotes pro-apoptotic response to genotoxic stress (Haydn 2014). Interestingly, R273H cells were sensitive to apoptosis while Y234C cells were insensitive to Doxorubicin treatment.

In summary, the neo-morphic phenotypes of the p53 missense mutants or the phenotypic heterogeneity observed across the p53 mutants can be attributed to altered DNA binding by mutated p53 transcription factors resulting in up or down regulation of target genes (canonical and non-canonical) which in turn affect cellular pathways.

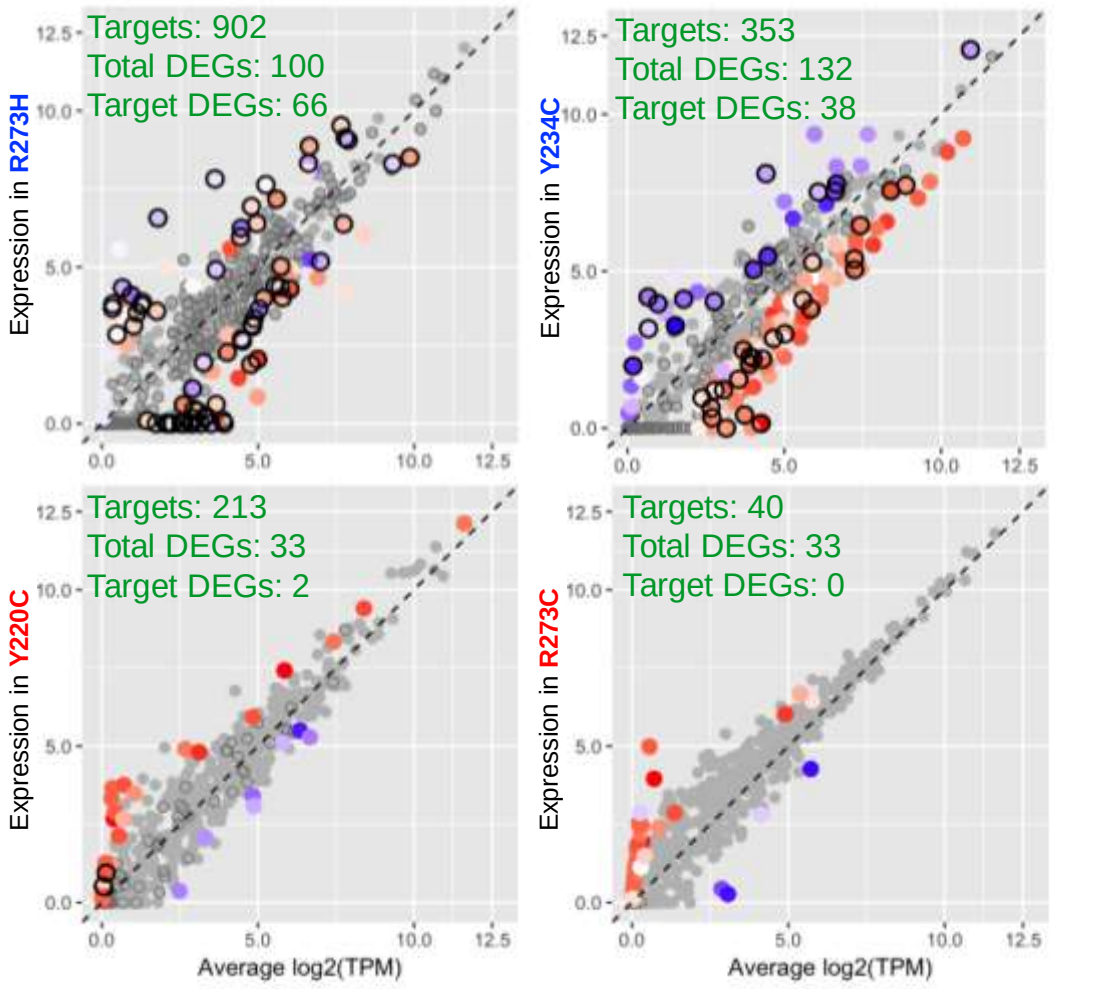


Figure 2.11. **Integrated analysis of ChIP Seq and RNA Seq**. Expression of 1,263 combined target genes of all mutant p53 proteins in each cell were compared to average expression in other 9 mutant p53-expressing cells. Color coded are differentially expressed genes (DEG, $P < 0.05$), the number of direct targets of each mutant, DEGs, and target genes among the

Conclusion

TP53 is the most frequently mutated gene occurring in a variety of sporadic human cancers and its mutation in the germline causes early onset of familial cancers in Li-Fraumeni syndrome illuminating p53 as a major tumor suppressor protein. While the tumor suppressors are inactivated by deletion or truncation mutations, the most common genetic aberration in *TP53* is the missense mutations that generate a full-length protein with a single altered amino acid at a crucial codon. The missense or point mutations concentrate at the DNA binding domain of the *TP53* transcription factor and affect its DNA sequence specific binding activity.

The p53 protein is active as a homo-tetramer but hetero-tetramerization of mutant with wildtype p53 monomers exhibits dominant negative effect over the wildtype p53. The gain-of-function (GOF) properties of the mutants in complete absence of wildtype allele due to loss of heterozygosity (LOH) at the *TP53* genetic locus contributes to the oncogenic characteristics of this mutant tumor suppressor. Thus, both the dominant negative effect and GOF impart 'neo-morphic' function to the mutant p53 protein.

In cancer tumors harboring *TP53* missense mutations, 60% show LOH but the rest 40% retain the wildtype *TP53* allele showing the relevance of dominant negative mechanism of the mutants to inhibit wildtype p53. Moreover, the knock-in mouse strains with *p53R172H* and *p53R270H* (equivalent to human hotspot missense mutations R175H and R273H) in combination with wild-type or null *TP53* allele show wide spectrum of tumors, tumor burden and metastasis compared to the *p53+/-* and *p53-/-* mice. The dominant negative effects vary by the allele. Carcinomas were common in *p53R270H* and osteosarcomas in *p53R172H* (Olive 2004). *p53R248Q* mutant HUPKI mice had rapid tumor development, decreased survival time and different tumor spectrum in comparison to *p53G245S* and *p53null* mice. Both *p53R248Q* and *p53G245S* showed higher diversity of sarcomas, more carcinomas and germ cell tumors than p53 null mice (Hanel 2013).

The mice studies support that the full-length p53 missense mutant proteins are not created equal and are rather heterogeneous (Mello 2013).

The TNBCs produce the most heterogeneous and aggressive tumors and 90% of them harbor mutations in the *TP53* gene. More than 100 different missense mutations are reported to occur along the length of *TP53* and it is likely that the mutant p53 proteins contribute to the tumor heterogeneity observed in the TNBCs. But how are these missense mutant p53 proteins different from one another and how do they contribute to tumor heterogeneity remains a mystery.

We selected a panel of ten most prevalent missense mutations of *TP53* occurring in breast tumor samples in the DBD of the protein. We expressed them in the non-transformed mammary epithelial cells, MCF 10A, to model the accumulation of mutant p53 in tumor cells and used them to study the phenotypic effect based on prominent hallmarks of cancer. The 10 different mutants produced heterogeneous phenotypes but based on the phenotype scores, some mutants were found to be more aggressive than the others. The DNA contact mutants R248W, R273C and the structural mutant Y220C were identified as the most aggressive based on their survival capacity in absence of critical growth factors, apoptosis and anoikis resistance, increased cell migration, invasion and formation of reverse polarized distorted mammospheres. In contrast, G245S mutant was the least aggressive with completely opposite phenotypes as compared to the aggressive mutants. Mutants R273H and Y234C were comparable to G245S. The rest of the p53 mutants formed a rainbow of phenotypes substantiating the fact that the different p53 missense mutant proteins are not functionally equal and these missense mutant proteins act more like oncogenes with ‘neo-morphic’ activities than loss-of-function tumor suppressors. We emphasize that our phenotypic results are reflective of the dominant negative effect of mutant p53 as the MCF 10A cells retain the endogenous wildtype p53. Such a system models the condition of early phase tumors that harbor *TP53* mutations in only one allele and eventually lose the remaining allele (deletion or mutation) by LOH during tumor progression.

Our phenotypic data correlates with a patient study data. In a study of 1,794 women with primary breast cancer, specific missense mutations at codon 179 and R248W were associated with worse prognosis. Breast cancer patients with missense mutation in DBD of p53 showed reduced survival compared to patients without the mutations and the 10-year mortality rates for DBD and non-DBD mutations were 73 and 44 (per 1,000 patients) respectively. Where missense mutation at codon 179 and R248W had reduced survival, G245S had better survival compared to any other missense mutation. Other mutation hotspots in breast cancer including R175H, R248Q, R273H, R273C, codon 163, 249 and 282 had mortality rates like other non-missense mutations in p53 (Olivier 2006). Our phenotype data on the p53 mutant proteins matches with the study except for the results of codon 179. This can be expected as the *TP53* mutations in patients appear in complex mutation background where thousands of other somatic mutations co-exist and the phenotypic outcome is the result of gene interactions. On the other hand, p53 mutations were the only mutations introduced in our model system.

Several other reported studies support our finding of grouping the R248W, R273C and Y220C as the most aggressive p53 mutants. The hotspot codon 248 is frequently mutated in luminal B, HER2 enriched and basal-like breast tumor subtypes. The other hotspot codons 175 and 273 are frequently mutated in basal-like subtype which is the most aggressive breast cancer subtype (Silwal-Pandit 2014). Moreover, in a patient cohort with locally advanced breast cancer treated with Doxorubicin monotherapy, some patients showed progressive disease during the treatment and harbored mutations in the L3 loop of the *TP53* gene (codon 273 and 248) that directly contacts DNA. The patients with mutations in the L2/L3 domain of p53 predicted poor survival and suggested the association of these specific p53 mutations with primary resistance to Doxorubicin therapy in breast cancer patients (Aas 1996). In a very recent study, mice with somatic mutation in *TP53* R245W (equivalent to human R248W) produced most aggressive and metastatic breast tumors and R172H (human R175H) mutated tumors were less aggressive. $Trp53^{R245W/+}$ mice developed breast tumors with 46% metastasis to lung and the liver whereas loss of heterozygosity (LOH) was needed for $Trp53^{R172H/-}$ mice to develop mammary tumors and

few of them disseminated to lungs. Early dissemination of mammary tumors was observed in R245W mice and thus, constant monitoring of human breast cancer patients harboring R248W mutations in primary tumors could prove beneficial. The mammary tumors from such mice models recapitulated the characteristics of human breast cancers making this an invaluable model to study the initiating event in breast cancer progression and the aggressive behavior of the *TP53* mutations (Zhang 2018).

Our study is a comprehensive resource for p53 missense mutant protein functional activities and transcriptomics. Compared to previous conventional research on p53 mutants, our study carefully integrated the cellular phenotypes based on the hallmarks of cancer with RNA seq and ChIP Seq data. We included 10 p53 missense mutants and an elaborate set of controls, the MCF 10A cells with endogenous wildtype p53, 10A cells overexpressing wildtype p53 and cells with knocked-down p53 (p53 null). To explore the underlying basis for the neo-morphic functions of missense mutant p53 proteins, we performed ChIP Seq on invasive mutants (R273C, Y220C) and non-invasive mutants (Y234C, R273H) with MCF 10A $p53^{OE}$ as the control. We used untreated cells and immunoprecipitated mutant p53 with a V5-tag. We found evidence that p53 missense mutant protein experiences altered DNA binding. The binding affinity to consensus p53 motif was low in all mutant p53 expressing cell lines compared to wildtype $p53^{OE}$. The altered DNA binding patterns of p53 mutants translated into the neo-morphic activities and the heterogeneous phenotypes. The RNA seq and the pathway analysis showed good correlation between the dysregulated pathways and the cancer related phenotypes across the mutant p53 expressing MCF 10A cell lines. We discovered that focal adhesion pathway was one of the highly dysregulated pathways in the invasive p53 mutants.

We also observed a close structure-function relationship of the mutant p53 proteins. The p53 hotspot mutant proteins in cancer preserve the overall structural scaffold but display different local structural changes that affect DNA binding and facilitate interactions with other proteins and provide the GOF properties. The mutant specific local structural changes refer to removal of DNA

contact point (by DNA contact mutants), conformational changes on the DNA binding surface and formation of surface cavities (by structural mutants). We observed that a subtle change in a single amino acid at a codon results in remarkable changes in the phenotype. Though the DNA contact mutations remove just a DNA contact residue without introducing large structural perturbations and preserving the architecture of the DNA-binding surface, replacement with a large hydrophobic side chain prevented sequence specific DNA binding. For example, R248W severely affects the hallmarks of cancer than R248Q and such a change could be attributed to the introduction of hydrophobic Tryptophan (R248W) instead of positively charged Arginine or polar and positively charged Glutamine (R248Q). Also, replacement of Arginine with small amino acid Cysteine (R273C) results in more aggressive behavior than R273H mutant where replacement with Histidine was comparable to Arginine. At codon 245, substitution of Glycine with Serine, doesn't severely impair the structure of p53 and that is what we observe in terms of the phenotypes generated by the mutant G245S whose phenotypes where similar to 10A cells with wildtype p53. The structural changes in G245S are much smaller and the information of the L3 loop is conserved.

Introduction of bulky Histidine at codon 175 causes structural distortions and disrupts the zinc-binding pocket of p53. Substitution of hydrophobic amino acid side chain Tyrosine with small Cysteine residue at Y163C, Y220C and Y234C created differential phenotypes with an aggressive phenotype associated with Y220C mutant that contains a large surface crevice on the mutant protein that perturbs the packing of the β-sandwich. The DNA contact mutant R273H has a subtle effect on the thermodynamic stability of the protein whereas the structural mutant R175H causes strong structural perturbation destabilizing the core domain. Such changes in the thermodynamic stability affect the DNA binding capacity of the mutant p53 (Joerger 2007). It is also reported that R273H has half-life similar to that of wildtype p53. Taking the structural studies on p53 mutant proteins into consideration, we find a good correlation between the deviation in structure and function (phenotypes) of the p53 missense mutant proteins though additional structural evidence is required to explain changes in function of mutant p53.

When mutant p53 proteins are over-expressed, such as the MCF 10A cells expressing mutant p53 in our study, formation of the mixed hetero-tetramers between mutant and wildtype p53 drastically reduces normal transactivation by wildtype p53. Ideally due to the co-existence of wild-type p53 with the mutant p53 in the MCF 10A cells, the phenotypes would be attributed to the dominant negative effect of the mutant p53 proteins but the overexpression of the p53 mutant proteins and the resultant phenotypes produced by them tilts more towards over-dominance or GOF properties of mutant p53 imparting 'neo-morphic' characteristics to the p53 mutant MCF 10A cells.

As the genetic background in the MCF 10A cells overexpressing the p53 missense mutant proteins is identical, our data highlights the neo-morphic activities of the p53 missense mutant proteins generating a spectrum of phenotypes based on the hallmarks of cancer supported by data at the molecular level from ChIP and RNA seq. The results may explain the phenotype heterogeneity observed in the TNBC tumors that contain different *TP53* mutations. Hence, identification of the exact mutation in *TP53* in human TNBC patients and its correlation with resistance to chemotherapy and poor prognosis may inform the course of therapy and stratify patients for effective clinical outcome.

References

Aas, Turid, Anne-Lise Børresen, Stephanie Geisler, Birgitte Smith-Sørensen, Hilde Johnsen, Jan E. Varhaug, Lars A. Akslen, and Per E. Lønning. "Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients." *Nature medicine* 2, no. 7 (1996): 811.

Aas, Turid, Anne-Lise Børresen, Stephanie Geisler, Birgitte Smith-Sørensen, Hilde Johnsen, Jan E. Varhaug, Lars A. Akslen, and Per E. Lønning. "Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients." *Nature medicine* 2, no. 7 (1996): 811.

Alexandrova, Evguenia M., Safia A. Mirza, Sulan Xu, Ramona Schulz-Heddergott, Natalia D. Marchenko, and Ute M. Moll. "p53 loss-of-heterozygosity is a necessary prerequisite for mutant p53 stabilization and gain-of-function in vivo." *Cell death & disease* 8, no. 3 (2017): e2661.

Ashcroft, Margaret, and Karen H. Vousden. "Regulation of p53 stability." *Oncogene* 18, no. 53 (1999): 7637.

Aubrey, Brandon J., Ana Janic, Yunshun Chen, Catherine Chang, Elizabeth C. Lieschke, Sarah T. Diepstraten, Andrew J. Kueh et al. "Mutant TRP53 exerts a target gene-selective dominant-negative effect to drive tumor development." *Genes & development* 32, no. 21-22 (2018): 1420-1429.

Bailey, Timothy L., Mikael Boden, Fabian A. Buske, Martin Frith, Charles E. Grant, Luca Clementi, Jingyuan Ren, Wilfred W. Li, and William S. Noble. "MEME SUITE: tools for motif discovery and searching." *Nucleic acids research* 37, no. suppl_2 (2009): W202-W208.

Billant, Olivier, Alice Léon, Solenn Le Guellec, Gaëlle Friocourt, Marc Blondel, and Cécile Voisset. "The dominant-negative interplay between p53, p63 and p73: A family affair." *Oncotarget* 7, no. 43 (2016): 69549.

Blandino, Giovanni, Arnold J. Levine, and Moshe Oren. "Mutant p53 gain of function: differential effects of different p53 mutants on resistance of cultured cells to chemotherapy." *Oncogene* 18, no. 2 (1999): 477.

Bode, Ann M., and Zigang Dong. "Post-translational modification of p53 in tumorigenesis." *Nature Reviews Cancer*4, no. 10 (2004): 793.

Brosh, Ran, and Varda Rotter. "When mutants gain new powers: news from the mutant p53 field." *Nature reviews cancer* 9, no. 10 (2009): 701.

Brosh, Ran, and Varda Rotter. "When mutants gain new powers: news from the mutant p53 field." *Nature reviews cancer* 9, no. 10 (2009): 701.

Cao, Zheng, Theodore Livas, and Natasha Kyprianou. "Anoikis and EMT: Lethal" liaisons" during cancer progression." *Critical Reviews™ in Oncogenesis* 21, no. 3-4 (2016).

De Vries, Annemieke, Elsa R. Flores, Barbara Miranda, Harn-Mei Hsieh, Conny Th M. van Oostrom, Julien Sage, and Tyler Jacks. "Targeted point mutations of p53 lead to dominant-negative inhibition of wild-type p53 function." *Proceedings of the National Academy of Sciences* 99, no. 5 (2002): 2948-2953.

Debnath, Jayanta, Senthil K. Muthuswamy, and Joan S. Brugge. "Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures." *Methods* 30, no. 3 (2003): 256-268.

Derksen, Patrick WB, Xiaoling Liu, Francis Saridin, Hanneke van der Gulden, John Zevenhoven, Bastiaan Evers, Judy R. van Beijnum et al. "Somatic inactivation of E-cadherin and p53 in mice leads to metastatic lobular mammary carcinoma through induction of anoikis resistance and angiogenesis." *Cancer cell* 10, no. 5 (2006): 437-449.

Di Agostino, Silvia, Sabrina Strano, Velia Emiliozzi, Valentina Zerbini, Marcella Mottolese, Ada Sacchi, Giovanni Blandino, and Giulia Piaggio. "Gain of function of mutant p53: the mutant p53/NF-Y protein complex reveals an aberrant transcriptional mechanism of cell cycle regulation." *Cancer cell* 10, no. 3 (2006): 191-202.

Dittmer, Dirk, Sibani Pati, Gerard Zambetti, Shelley Chu, Angelika K. Teresky, Mary Moore, Cathy Finlay, and Arnold J. Levine. "Gain of function mutations in p53." *Nature genetics* 4, no. 1 (1993): 42.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. "STAR: ultrafast universal RNA-seq aligner." *Bioinformatics* 29, no. 1 (2013): 15-21.

Donehower, Lawrence A., Michele Harvey, Betty L. Slagle, Mark J. McArthur, Charles A. Montgomery Jr, Janet S. Butel, and Allan Bradley. "Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours." *Nature* 356, no. 6366 (1992): 215.

Freed-Pastor, William A., and Carol Prives. "Mutant p53: one name, many proteins." *Genes & development* 26, no. 12 (2012): 1268-1286.

Gilmore, A. P. "Anoikis." (2005): 1473.

Goh, Amanda M., Cynthia R. Coffill, and David P. Lane. "The role of mutant p53 in human cancer." *The Journal of pathology*223, no. 2 (2011): 116-126.

Hanahan, Douglas, and Robert A. Weinberg. "Hallmarks of cancer: the next generation." *cell* 144, no. 5 (2011): 646-674.

Hanel, W., N. Marchenko, S. Xu, S. Xiaofeng Yu, W. Weng, and U. Moll. "Two hot spot mutant p53 mouse models display differential gain of function in tumorigenesis." *Cell death and differentiation* 20, no. 7 (2013): 898.

Harvey, Michele, Hannes Vogel, Danna Morris, Allan Bradley, Alan Bernstein, and Lawrence A. Donehower. "A mutant p53 transgene accelerates tumour development in

heterozygous but not nullizygous p53–deficient mice." *Nature genetics* 9, no. 3 (1995): 305.

Heinz, Sven, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh, and Christopher K. Glass. "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." *Molecular cell* 38, no. 4 (2010): 576-589.

Huang, Wenhuan, Cheng Chen, Zhongheng Liang, Junlu Qiu, Xinxin Li, Xiang Hu, Shuanglin Xiang, Xiaofeng Ding, and Jian Zhang. "AP-2α inhibits hepatocellular carcinoma cell growth and migration." *International journal of oncology* 48, no. 3 (2016): 1125-1134.
Ikeda, Osamu, Yuichi Sekine, Akihiro Mizushima, Misa Nakasuji, Yuto Miyasaka, Chikako Yamamoto, Ryuta Muromoto et al. "Interactions of STAP-2 with Brk and STAT3 participate in cell growth of human breast cancer cells." *Journal of Biological Chemistry* 285, no. 49 (2010): 38093-38103.

Ilić, Duško, Eduardo AC Almeida, David D. Schlaepfer, Paul Dazin, Shinichi Aizawa, and Caroline H. Damsky. "Extracellular matrix survival signals transduced by focal adhesion kinase suppress p53-mediated apoptosis." *The Journal of cell biology* 143, no. 2 (1998): 547-560.

Jacks, Tyler, Lee Remington, Bart O. Williams, Earlene M. Schmitt, Schlomit Halachmi, Roderick T. Bronson, and Robert A. Weinberg. "Tumor spectrum analysis in p53-mutant mice." *Current biology* 4, no. 1 (1994): 1-7.

Joerger, A. C., and A. R. Fersht. "Structure–function–rescue: the diverse nature of common p53 cancer mutants." *Oncogene* 26, no. 15 (2007): 2226.

Kalluri, Raghu, and Robert A. Weinberg. "The basics of epithelial-mesenchymal transition." *The Journal of clinical investigation* 119, no. 6 (2009): 1420-1428.

Kenny, Paraic A., and Mina J. Bissell. "Tumor reversion: correction of malignant behavior by microenvironmental cues." *International journal of cancer* 107, no. 5 (2003): 688-695.

Kim, Jongdoo, Seunghee Bae, Sungkwan An, Jong Kuk Park, Eun Mi Kim, Sang-Gu Hwang, Wun-Jae Kim, and Hong-Duck Um. "Cooperative actions of p21WAF1 and p53 induce Slug protein degradation and suppress cell invasion." *EMBO reports* 15, no. 10 (2014): 1062-1068.

Kim, Michael P., and Guillermina Lozano. "Mutant p53 partners in crime." *Cell death and differentiation* 25, no. 1 (2018): 161.

Lang, Gene A., Tomoo Iwakuma, Young-Ah Suh, Geng Liu, V. Ashutosh Rao, John M. Parant, Yasmine A. Valentin-Vega et al. "Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome." *Cell* 119, no. 6 (2004): 861-872.

Langmead, Ben, and Steven L. Salzberg. "Fast gapped-read alignment with Bowtie 2." *Nature methods* 9, no. 4 (2012): 357.

Levine, Arnold J., Jamil Momand, and Cathy A. Finlay. "The p53 tumour suppressor gene." *nature* 351, no. 6326 (1991): 453.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. "The sequence alignment/map format and SAMtools." *Bioinformatics* 25, no. 16 (2009): 2078-2079.

Li, Jie, Lixin Yang, Shikha Gaur, Keqiang Zhang, Xiwei Wu, Yate-Ching Yuan, Hongzhi Li, Shuya Hu, Yaguang Weng, and Yun Yen. "Mutants TP 53 p. R273H and p. R273C but not p. R273G Enhance Cancer Cell Malignancy." *Human mutation*35, no. 5 (2014): 575-584.

Liu, D. P., Hoseok Song, and Yang Xu. "A common gain of function of p53 cancer mutants in inducing genetic instability." *Oncogene* 29, no. 7 (2010): 949.

Liu, Geng, Timothy J. McDonnell, Roberto Montes de Oca Luna, Mini Kapoor, Betsy Mims, Adel K. El-Naggar, and Guillermina Lozano. "High metastatic potential in mice inheriting a targeted p53 missense mutation." *Proceedings of the National Academy of Sciences* 97, no. 8 (2000): 4174-4179.

Liu, Xiaoling, Henne Holstege, Hanneke van der Gulden, Marcelle Treur-Mulder, John Zevenhoven, Arno Velds, Ron M. Kerkhoven et al. "Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1-mutated basal-like breast cancer." *Proceedings of the National Academy of Sciences* 104, no. 29 (2007): 12111-12116.

Lowe, Scott W., Stephan Bodis, Andrea McClatchey, Lee Remington, H. Earl Ruley, David E. Fisher, David E. Housman, and Tyler Jacks. "p53 status and the efficacy of cancer therapy in vivo." *Science* 266, no. 5186 (1994): 807-810.

Marchese, Stephanie, and Elisabete Silva. "Disruption of 3D MCF-12A breast cell cultures by estrogens–an in vitro model for ER-mediated changes indicative of hormonal carcinogenesis." *PloS one* 7, no. 10 (2012): e45767.

Mello, Stephano Spano, and Laura D. Attardi. "Not all p53 gain-of-function mutants are created equal." (2013): 855.

Muller, Patricia AJ, and Karen H. Vousden. "p53 mutations in cancer." *Nature cell biology* 15, no. 1 (2013): 2.

Nishida, Kozo, Keiichiro Ono, Shigehiko Kanaya, and Koichi Takahashi. "KEGGscape: a Cytoscape app for pathway data integration." *F1000Research* 3 (2014).

Olive, Kenneth P., David A. Tuveson, Zachary C. Ruhe, Bob Yin, Nicholas A. Willis, Roderick T. Bronson, Denise Crowley, and Tyler Jacks. "Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome." *Cell* 119, no. 6 (2004): 847-860.

Olive, Kenneth P., David A. Tuveson, Zachary C. Ruhe, Bob Yin, Nicholas A. Willis, Roderick T. Bronson, Denise Crowley, and Tyler Jacks. "Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome." *Cell* 119, no. 6 (2004): 847-860.

Olivier, Magali, Anita Langer, Patrizia Carrieri, Jonas Bergh, Sigrid Klaar, Jorunn Eyfjord, Charles Theillet et al. "The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer." *Clinical cancer research* 12, no. 4 (2006): 1157-1167.

Oren, Moshe, and Varda Rotter. "Mutant p53 gain-of-function in cancer." *Cold Spring Harbor perspectives in biology* 2, no. 2 (2010): a001107.

Powell, Emily, David Piwnica-Worms, and Helen Piwnica-Worms. "Contribution of p53 to metastasis." *Cancer discovery*4, no. 4 (2014): 405-414.

Rivlin, Noa, Ran Brosh, Moshe Oren, and Varda Rotter. "Mutations in the p53 tumor suppressor gene: important milestones at the various steps of tumorigenesis." *Genes & cancer* 2, no. 4 (2011): 466-474.

Roger, Lauréline, Laurent Jullien, Véronique Gire, and Pierre Roux. "Gain of oncogenic function of p53 mutants regulates E-cadherin expression uncoupled from cell invasion in colon cancer cells." *J Cell Sci* 123, no. 8 (2010): 1295-1305.

Santoro, Angela, Thalia Vlachou, Lucilla Luzi, Giorgio Melloni, Luca Mazzarella, Errico D'Elia, Xieraili Aobuli et al. "p53 Loss in Breast Cancer Leads to Myc Activation, Increased Cell Plasticity, and Expression of a Mitotic Signature with Prognostic Value." *Cell reports* 26, no. 3 (2019): 624-638.

Sayols, Sergi, Denise Scherzinger, and Holger Klein. "dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data." *BMC bioinformatics* 17, no. 1 (2016): 428.

Sekine, Yuichi, Osamu Ikeda, Satoshi Tsuji, Chikako Yamamoto, Ryuta Muromoto, Asuka Nanbo, Kenji Oritani, Akihiko Yoshimura, and Tadashi Matsuda. "Signal-

transducing adaptor protein-2 regulates stromal cell-derived factor-1α-induced chemotaxis in T cells." *The Journal of Immunology*183, no. 12 (2009): 7966-7974.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. "Cytoscape: a software environment for integrated models of biomolecular interaction networks." *Genome research* 13, no. 11 (2003): 2498-2504.

Silwal-Pandit, Laxmi, Hans Kristian Moen Vollan, Suet-Feung Chin, Oscar M. Rueda, Steven McKinney, Tomo Osako, David A. Quigley et al. "TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance." *Clinical Cancer Research* 20, no. 13 (2014): 3569-3580.

Song, Hoseok, Monica Hollstein, and Yang Xu. "p53 gain-of-function cancer mutants induce genetic instability by inactivating ATM." *Nature cell biology* 9, no. 5 (2007): 573.

Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich et al. "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." *Proceedings of the National Academy of Sciences* 102, no. 43 (2005): 15545-15550.

Tait, Larry, Herbert D. Soule, and Jose Russo. "Ultrastructural and immunocytochemical characterization of an immortalized human breast epithelial cell line, MCF-10." *Cancer research*50, no. 18 (1990): 6087-6094.

Tan, B. S., K. H. Tiong, H. L. Choo, F. Fei-Lei Chung, L. W. Hii, S. H. Tan, I. KS Yap et al. "Mutant p53-R273H mediates cancer cell survival and anoikis resistance through AKT-dependent suppression of BCL2-modifying factor (BMF)." *Cell death & disease* 6, no. 7 (2015): e1826.

Vaughan, Catherine A., Swati P. Deb, Sumitra Deb, and Brad Windle. "Preferred binding of gain-of-function mutant p53 to bidirectional promoters with coordinated binding of ETS1 and GABPA to multiple binding sites." *Oncotarget* 5, no. 2 (2014): 417.

Ventura, Andrea, David G. Kirsch, Margaret E. McLaughlin, David A. Tuveson, Jan Grimm, Laura Lintault, Jamie Newman, Elizabeth E. Reczek, Ralph Weissleder, and Tyler Jacks. "Restoration of p53 function leads to tumour regression in vivo." *Nature* 445, no. 7128 (2007): 661.

Wang, Jing, Gregory J. Hannon, and David H. Beach. "Cell biology: risky immortalization by telomerase." *Nature* 405, no. 6788 (2000): 755.

Wang, Shu-Ping, Wen-Lung Wang, Yih-Leong Chang, Chen-Tu Wu, Yu-Chih Chao, Shih-Han Kao, Ang Yuan et al. "p53 controls cancer cell invasion by inducing the MDM2-mediated degradation of Slug." *Nature cell biology* 11, no. 6 (2009): 694.

Wijnhoven, Susan WP, Ewoud N. Speksnijder, Xiaoling Liu, Edwin Zwart, Rudolf B. Beems, Esther M. Hoogervorst, Mirjam M. Schaap et al. "Dominant-negative but not gain-of-function effects of a p53. R270H mutation in mouse epithelium tissue after DNA damage." *Cancer research* 67, no. 10 (2007): 4648-4656.

Willis, Amy, Eun Joo Jung, Therese Wakefield, and Xinbin Chen. "Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes." *Oncogene* 23, no. 13 (2004): 2330.

Yusuf, Rita, and Krystyna Frenkel. "Morphologic transformation of human breast epithelial cells MCF-10A: dependence on an oxidative microenvironment and estrogen/epidermal growth factor receptors." *Cancer Cell International* 10, no. 1 (2010): 30.

Zhang, Yanhong, Wensheng Yan, and Xinbin Chen. "Mutant p53 disrupts MCF-10A cell polarity in three-dimensional culture via epithelial-to-mesenchymal transitions." *Journal of Biological Chemistry* 286, no. 18 (2011): 16218-16228.

Zhang, Yong, Tao Liu, Clifford A. Meyer, Jérôme Eeckhoute, David S. Johnson, Bradley E. Bernstein, Chad Nusbaum et al. "Model-based analysis of ChIP-Seq (MACS)." *Genome biology* 9, no. 9 (2008): R137.

Zhang, Yun, Shunbin Xiong, Bin Liu, Vinod Pant, Francis Celii, Gilda Chau, Ana C. Elizondo-Fraire et al. "Somatic Trp53 mutations differentially drive breast cancer and evolution of metastases." *Nature communications* 9, no. 1 (2018): 3953.

Zhao, Yahui, Xiaoding Hu, Li Wei, Dan Song, Juanjuan Wang, Lifang You, Hexige Saiyin et al. "PARP10 suppresses tumor metastasis through regulation of Aurora A activity." *Oncogene*(2018): 1.

Zheng, Tongsen, Jiabei Wang, Yuhan Zhao, Cen Zhang, Meihua Lin, Xiaowen Wang, Haiyang Yu, Lianxin Liu, Zhaohui Feng, and Wenwei Hu. "Spliced MDM2 isoforms promote mutant p53 accumulation and gain-of-function in tumorigenesis." *Nature communications* 4 (2013): 2996.

Zhu, Jiajun, Morgan A. Sammons, Greg Donahue, Zhixun Dou, Masoud Vedadi, Matthäus Getlik, Dalia Barsyte-Lovejoy et al. "Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth." *Nature* 525, no. 7568 (2015): 206.

Zilfou, Jack T., and Scott W. Lowe. "Tumor suppressive functions of p53." *Cold Spring Harbor perspectives in biology*1, no. 5 (2009): a001883.

CHAPTER 3

PTEN AS A CO-DRIVER OF MISSENSE MUTANT P53 TO DRIVE CELL INVASION: A PILOT STUDY

Introduction

Next generation sequencing (NGS) has identified millions of somatic mutations in the cells derived from cancer tumors. An important challenge is to distinguish the driver mutations in a sea of neutral passenger mutations that accelerate cancer development. For the identification of candidate driver genes frequency-based and function-based approaches have been developed.

The frequency-based approach relies on candidate driver genes mutated in greater proportion of cancer sample than expected from the background mutation rate (Pon 2015). Whether a candidate driver is a tumor suppressor, or an oncogene is determined by assessing the relative frequency and the distribution of missense, non-sense and frameshift mutations. Recurrent mutations are characteristic of driver mutations. But infrequent mutations in driver genes are abundant in the mutational landscape of cancer tumors and present in <5% of samples. The frequency-based approaches often overlook the infrequently mutated genes as drivers and a huge sample size is needed to identify these infrequent drivers by bulk sequencing method.

The function-based approach distinguishes the candidate drivers from the passengers by their tendency to have a selective advantage on protein function compared to the non-advantageous neutral passenger mutations. The functional impact of the candidate driver mutations can be predicted from the evolutionary conservation of the protein, its structural features and changes in amino acid sequence affecting the 3D structure of the protein. Other methods under this approach include studying different pathways, cell phenotypes and transcriptome data. Moreover, this approach can directly distinguish the candidate driver genes as oncogenes or tumor

suppressors. Function-based approaches are the methods of choice to study intra and inter-tumor heterogeneity and develop precision targeted treatments.

The breast cancer tumors, specifically the TNBC tumors are highly heterogeneous and >88% of the samples contain mutations in the *TP53* tumor suppressor establishing it as one of the driver genes. But it exists in a sea of co-mutated genes that include the passenger mutations and the infrequent drivers which we term as the 'co-drivers'. We hypothesize that the observed heterogeneity in TNBC tumors is a result of functional cooperation between the neo-morphic p53 mutants and distinct sets of co-existing somatic mutations or the co-drivers. To identify the possible co-drivers of different *TP53* missense mutations we tried to establish a function-based experimental pipeline. To determine the workability of the pipeline, we performed a pilot study by deleting *PTEN* in cells expressing missense mutant p53. Deletion mutation in *PTEN* is a known co-driver of mutant *TP53* where they cooperate to drive increased cell invasion.

Aim

Establishing an entire pipeline to identify co-drivers of missense mutant p53 by performing a pilot study focusing on cell invasion phenotype with the known co-driver PTEN. The pipeline required optimization of parameter such as gene deletion, phenotype-based enrichment of cells with targeted gene deletion, confirmation of gene deletion by NGS and validation of the phenotype change.

Approach

PTEN deletion is a known co-driver of mutant p53 in Triple Negative Breast Cancer (TNBC)

Triple negative breast cancer (TNBC) is the most aggressive subtype that represents about 16% of all breast cancer tumors. These tumors lack the expression of estrogen, progesterone and

HER2 receptors and hence, are called triple negative. RNA and deep sequencing measurements show that there is much variation in the clonal frequencies of somatic mutations in these tumors. Somatic mutations in *TP53, PTEN* and *PIK3CA* are clonally dominant, and mutations in cytoskeletal and motility related genes occur at lower frequencies in TNBC (Shah 2012). The TCGA confirms that loss of *RB1* and *BRCA1* along with mutations in *TP53* and *PIK3CA* are TNBC or basal-like features.

Concomitant deletion of *PTEN* and *TP53* in mice increased the formation of claudin-low triple negative breast cancer with activated AKT signaling and mesenchymal morphology compared to tumors with only *PTEN* or *TP53* mutations (Liu 2014). In another study, knock-down of both p53 and PTEN proteins in human mammary epithelial cells (HNMECs) and MCF 10As resulted in metastatic tumors in mice whose gene expression profiles mimicked the profile of the basal/claudin-low breast cancer tumor subtypes of TNBC. The simultaneous knock-down also increased sphere forming ability and cell motility in vitro and led to formation of tumors in NOD/SCID mice that metastasized to lungs, while single deletion of either *TP53* or *PTEN* failed to produce such results (Kim 2015). Thus, mutations in both *TP53* and *PTEN* work synergistically towards an aggressive cellular phenotype but are not enough for full malignant transformation of cells as seen in patient tumor samples.

We used the MCF 10A cell-lines expressing mutant p53 with either knocked-down PTEN by use of shRNA or deleted *PTEN* gene by CRISPR-Cas9 editing as the control cell-lines to develop the pipeline.

PTEN as a tumor suppressor

The PI3K pathway is a central biochemical pathway that regulates cellular processes like cell metabolism, survival, proliferation and migration. The pathway is triggered by signaling through receptor tyrosine kinases (RTKs) or G-protein coupled receptors (GPCRs), which results in the

activation of the PI3 kinases (PI3K). PI3K then phosphorylates phosphatidylinositol-4,5-bisphosphate (PIP_2) into the signaling molecule phosphatidylinositol-3,4,5-triphosphate (PIP_3), which in turn triggers numerous downstream targets. Counterposing the activation of the pathway, the lipid phosphatase PTEN reverses this by dephosphorylating PIP_3 back into PIP_2 thereby negatively regulating the PI3K pathway (Fig 3.1).

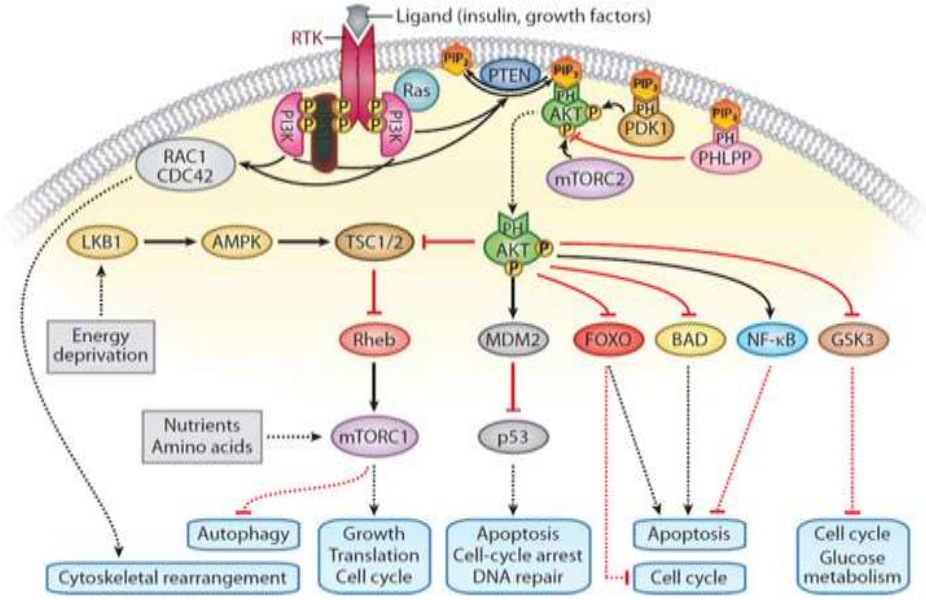


Figure 3.1. **The phosphatidylinositol 3-kinase (PI3K) signaling pathway.** Image adapted from (Chalhoub 2009).

Activation of the PI3K pathway drives cell proliferation. An important downstream effector of PIP_3 is AKT, a Ser/Thr kinase. Phosphorylation of transcription factor FOXO by AKT increases cell proliferation and survival. AKT directed phosphorylation of TSC2 activates mTORC1 that upregulates protein synthesis. AKT mediated phosphorylation of several proteins regulate cell cycle progression, survival and cell motility. In human cancers catalytic unit of PI3K, p110α, and *PTEN* are frequently mutated increasing the activity of the PI3K signaling pathway (Chalhoub 2009). Activation of PI3K-AKT-mTORC1 pathway influences metabolic re-programming of cancer cells and sustains their growth and proliferation via macromolecule (protein, lipid, nucleotide)

biosynthesis. Not surprisingly, activating mutations of the key effectors (PI3K, AKT) behave as oncogenes, whereas negative regulators, such as *PTEN*, behave as tumor suppressors.

Beyond opposing the PI3K pathway, PTEN protein activity also regulates Rho GTPases RAC1 and CDC42, thus reducing cell motility and maintain polarity. Dephosphorylation of FAK by *PTEN* decreases migration. Hence, *PTEN* deficient cancer cells lose cell polarity and show increased cell migration, i.e., become more mesenchymal. As a tumor suppressor gene deletions and somatic mutations in the *PTEN* (Phosphatase and tensin homologue deleted on chromosome 10) are very common in cancers of prostate, endometrium and glioblastoma multiforme.

Structure of PTEN

The crystal structure of *PTEN* reveals a N-terminal phosphatase domain and a C-terminal C2 domain (Lee 1999). The C2 domain is essential for phospholipid membrane binding and mutations in this region reduce membrane affinity and ability to suppress tumors (Fig 3.2). Most mutations occurring in *PTEN* are deletions and only a fraction of them are missense mutations. The missense mutation in *PTEN* occurs in the phosphatase domain and impairs the catalytic activity of PTEN protein in human cancers and Cowden syndrome. The C-terminal tail of the protein contains sites for phosphorylation that regulates PTEN stability, activity and its recruitment to the membrane (Georgescu 1999).

The prior knowledge of the protein domains is helpful while designing sgRNAs for gene editing by CRISPR-Ca9 method as described in the following section (*inactivation of PTEN: Use of the CRISPR-Cas9 system for gene editing).*

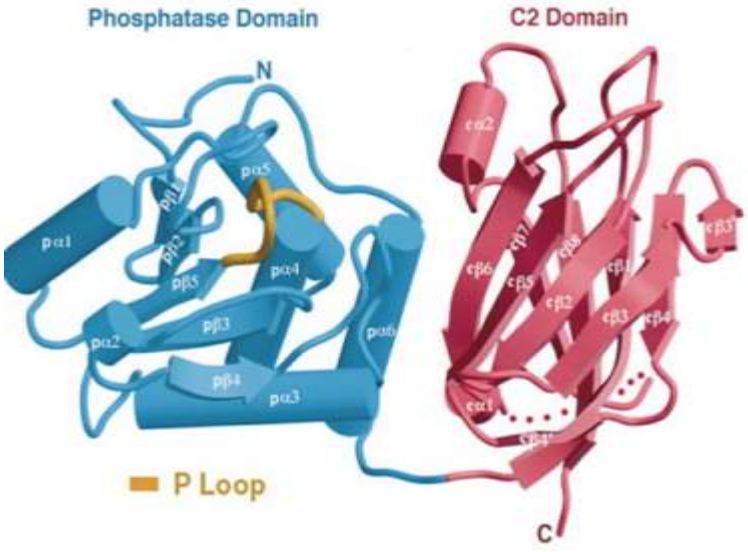


Figure 3.2. **The crystal structure of the *PTEN* tumor suppressor protein.** Image taken from (Lee 1999).

The two important tumor suppressor genes PTEN and TP53 and their interdependence

PTEN and *TP53* remain the most commonly inactivated or mutated genes in human cancers. Moreover, there is a great deal of functional and regulatory interdependence between the tumor suppressors. Analysis of the human *PTEN* genomic locus show the presence of a p53 DNA binding element at the PTEN promoter region and this element is essential for inducible transactivation of PTEN by p53. PTEN mRNA and protein levels rise with the induction of wildtype p53 and in turn PTEN is needed for p53 mediated cell apoptosis (Stambolic 2001). PTEN regulates the stability of p53 in a phosphatase-dependent manner by inhibiting Akt-Mdm2 pathway (Mayo 2002) where PTEN inhibits activation of Akt that fails to phosphorylate Mdm2 and restricts Mdm2 to the cytoplasm. The absence of Mdm2 in the nucleus stabilizes p53 protein levels by inhibiting ubiquitination of p53. In the phosphatase-independent manner PTEN forms a nuclear complex with p300 and keeps p53 acetylated, and the acetylated p53 forms tetramers and interacts with PTEN (Li 2006).

Inactivation of PTEN: PTEN protein knock-down by shRNA

RNA interference or RNAi is the most extensively used approach to disrupt gene function in mammalian cells. It is achieved by using short-interfering RNAs (siRNAs) or the short-hairpin RNAs (shRNAs). The dsRNA is introduced into a cell in an expression vector. Drosha/ DGCR8 complex process primary transcripts and form pre-shRNAs in the nucleus. Pre-shRNAs are loaded onto Dicer/TRBP/PCT complex in the cytoplasm to get processed into mature shRNA. The matured shRNA is finally loaded into the RNA induced silencing complex (RISC) that promotes the degradation of complementary target mRNA. RNAi targets mRNA and the subsequently reduces target protein levels (Rao 2009, Boettcher 2015). The RNAi silencing machinery is present in all mammalian cells and no prior genetic manipulation of the target cell is required. A direct siRNA or shRNA transfection results in loss-of-function phenotype.

Inactivation of PTEN: Use of the CRISPR-Cas9 system for gene editing

Clustered regularly interspaced short palindromic repeats (CRISPR) is a microbial adaptive immune system by which the host recognizes pathogenic DNA sequences and targets them for cleavage. This system has been adapted for use in the eukaryotic cells for genome engineering using a specific 20 nt targeting sequence as the guide RNA (gRNA). Such precise editing of genomic loci elucidates the function of genetic elements and genetic variations. The RNA guided CRISPR-Cas9 system has the guide RNA that base-pairs with the target DNA and the Cas9 endonuclease creates the double stand breaks (DSBs) after proper positioning (Fig 3.3). This system is easy to design, is quite specific and efficient, and can be used for high-throughput and multiplexed gene editing in various cell types and organisms.

DSB by Cas9 at the target locus may repair under the error-prone non-homologous end joining (NHEJ) DNA repair or high-fidelity homology directed repair (HDR). In absence of the repair

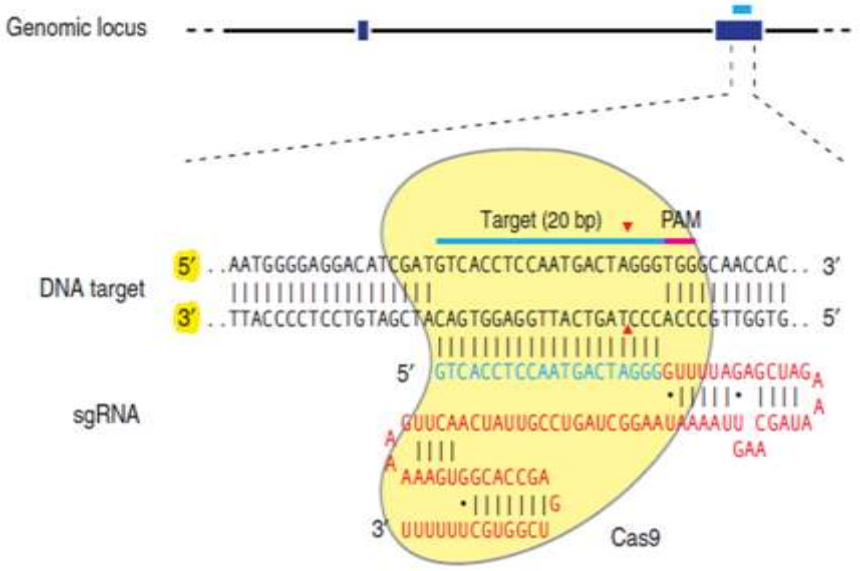


Figure 3.3. **The RNA guided Cas9 (CRISPR-Cas) system.** The 20 nt (blue) guide RNA with the (red) scaffold targets the Cas9 endonuclease (yellow) to the genomic loci. The DSB occurs at the red arrows. Adapted from (Ran 2013).

template, NHEJ leaves random insertions/ deletions (indels). Indels in the exons often cause frameshifts that lead to premature stop codons. Multiple DSBs can result in larger deletions at the targeted loci (Fig 3.4). The CRISPR-Cas9 system is reconstituted in the mammalian cells by expressing human codon-optimized Cas9 enzyme and the gRNA components. This Cas9 endonuclease can then be targeted specifically to any part of the genome preceding the PAM (Protospacer adjacent motif) sequence just by programming the 20 nt gRNA sequence with matching sequences (Ran 2013, Canver 2014). This CRISPR-Cas9 system is used mostly for gene knock-out and for several other purposes like gene repression or activation, specific genomic loci imaging and purification and for genome-wide screens.

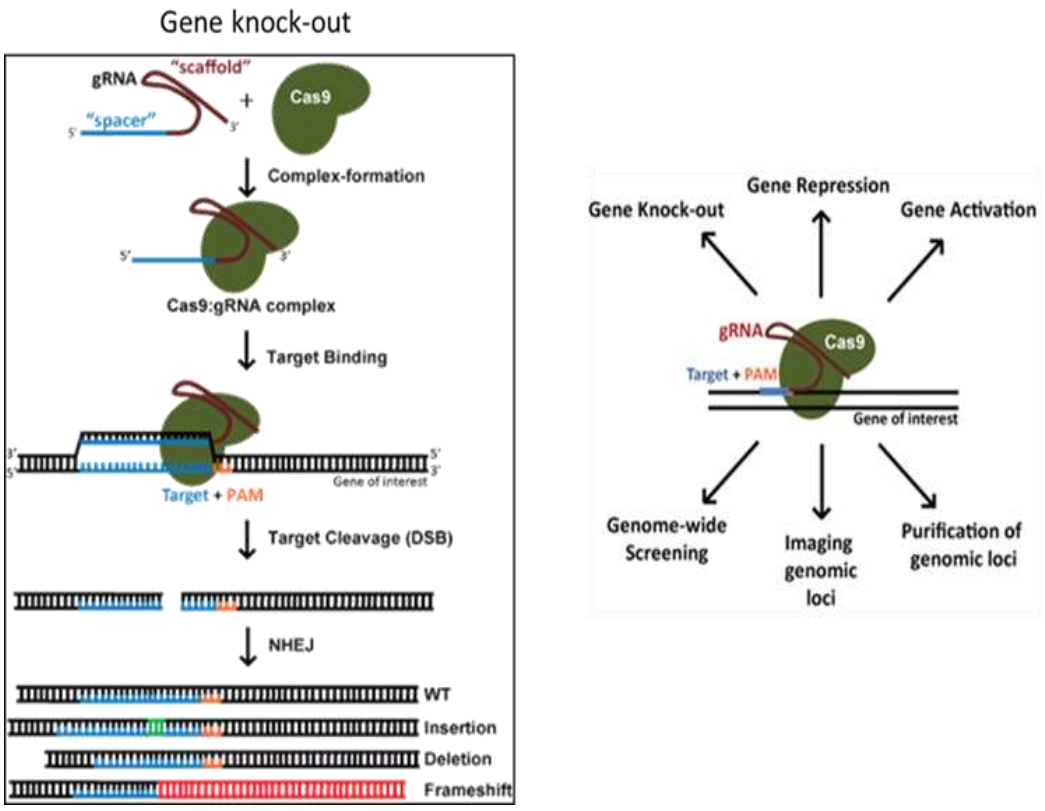


Figure 3.4. **Gene deletion by the CRISPR-Cas9 system and its other uses.** Figures adapted from www.addgene.org.

The shRNA causes mRNA cleavage and degradation resulting in deletion of protein. But RNAi doesn't always mirror the complete loss-of-function that results from gene mutation. RNAi is efficient in causing a knock-down than a knock-out. The ability to directly edit the genome and create gene knock-out makes the CRISPR-Cas9 a method of choice when mimicking somatic mutations under physiological conditions. Gene-knockout leads to complete depletion of functional protein.

Method

Downregulation of PTEN in MCF 10A cell lines using shRNA

The PTEN shRNA containing plasmids were obtained from our internal plasmid repository DNASU. After plasmid purification, the lentivirus was prepared in 6-well plates as following. On Day 1, 1.5 million 293Lx cells were plated in each well in 2 ml media. On Day 2 a transfection master mix was prepared containing mix 1 and mix 2. Mix 1 of total 150 ul contained diluted Fugene HD in serum free DMEM media (Fugene HD ul: total ug of DNA = 6: 1). Mix 2 had packaging plasmids 500 ng VSVG, 2500 ng Gag-pol and 2500 ng of lentivector containing the shRNA in serum free DMEM media making a total of 150 ul. Mix 1 and mix 2 were combined and incubated at room temperature (RT) for 20 to 30 min. After incubation, 150 ul of the master mix was dispensed into each of 2 wells. The volume of the master mix and the concentration of the plasmids was increased according to the number of wells transfected. The 6-well plate was spun at 2250 rpm for 30 min and placed in the incubator at 37^{o}C with 5% CO_2. 1% Wescodyne was used for disposing biohazard waste at every step. On Day 3, 1 ml media was removed and replaced with 2 ml fresh media in the morning and in the evening 2.5 ml media was removed and replaced with 5.5 ml fresh media in each well. On Day 4, 5.5 ml media was collected in a 50 ml conical tube and stored at 4^{o}C and 3.5 ml fresh media was put into the wells. On the last day, Day 5, 4 ml media was collected from each well in the same tube. The media was added into each well cautiously at a very slow speed not to disturb the 293Lx cell monolayer.

For the transduction of MCF 10A p53 mutant cell lines 80,000 cells/ well were plated in a 6-well plate. After 24 hours 1 ml media was removed and polybrene (8ug/ml) was added to each well which was followed by the addition of the lentivirus. On the next day the media in the plates was replaced by fresh media. The selection reagent, in this case puromycin (0.7 ug/ml), was added to the media on the following day and cells were selected for 72 hrs. As controls, we had un-transduced cells with and without puromycin. The un-transduced cells in puromycin were dead by

72 hours whereas the un-transduced cells without puromycin remained confluent. The cells were selected for at least 7 to 10 days in puromycin before using them for experiments. Cells were passaged at regular intervals and frozen stocks were also prepared.

Cell lysate preparation and western blot

100,000 cells/well were plated in 6-well plates for the cell lysate preparation. Cell lysates were made using the RIPA lysis buffer containing 50mM Tris, 150mM NaCl, 1% IGEPAL CA-630, 0.1% sodium azide (Sigma #S8032), cOmplete mini protease inhibitor (Sigma #04693124001), 200 µM Sodium fluoride and 200 µM of Sodium orthovanadate. After addition of 500 ul lysis buffer to each well, cells were scraped, incubated on ice for 10 minutes and centrifuged to collect the clear cell lysate. Lysates were quantified using the Pierce BCA kit (cat # 23225). Samples were run on 4-20% TGX gels from BioRad (#567-1093) and blotted onto 0.45µm PVDF membrane (GE Healthcare #10600023) using the BioRad semi-dry transfer system. Primary antibody for PTEN (CST catalogue no. 9188) and p53 (Sigma catalogue no. P6874) were diluted to 1:1000 and secondary HRP-linked anti-mouse or anti-rabbit antibody used to visualize the protein bands. The blots were imaged on the Fluorchem FC3 from AlphaInnotech.

Transwell cell invasion assay

Cells (70-80% confluent) were trypsinized and used for the transwell invasion assay. Before the experiment, the 24 well Matrigel coated invasion plate (Corning cat # 354480) was thawed with 200 ul serum free media in each well for rehydrating the Matrigel layer at 37^{o}C for 1 hr. 2.5 x 10^{5} cells in 200ul serum free media were seeded in each Matrigel coated insert. 750 ul of serum containing media was added to the bottom wells. The plate was incubated for 24 hrs. After incubation, cells were fixed with 4% paraformaldehyde for 2 min at RT, permeabilized by adding 500ul of 100% cold methanol (20 min at RT) and stained with 0.5% crystal violet for 15 min. The inserts were washed with 1X PBS 3 times. A cotton swab was used to remove any non-invasive

cells from inside of the insert. The invasive cells were imaged under the microscope at 4X and 10X magnification.

Designing and cloning the sgRNAs for PTEN in lentiCRISPR v2 plasmid

The sgRNAs targeting *PTEN* were designed using the online design tool from Feng Zhang's lab page and from the genetic perturbation platform of the BROAD Institute (Appendix A).

To clone the sgRNA sequences targeting *PTEN* into lentiCRISPR v2 plasmid, two oligos were synthesized for each sgRNA. After the vector digestion with BsmBI, the plasmids had an overhang that matched with that on the synthesized oligos. For the cloning of the appropriate sgRNAs into the lentiCRISPR v2 plasmid, the following steps were performed (Table 3.1).
1) Digestion of lentiCRISPR v2 plasmid. The vector was either digested and dephosphorylated or simply digested with BsmBI restriction enzyme without dephosphorylation. 2) Gel extraction of cut plasmid. BsmBI digestion resulted in a 2 kb filler fragment and a larger band. The digested products were checked on 1% agarose gel run for 27 min at 120 V. For cloning, only the larger band was extracted from the 1% GTG agarose gel (with gelstar) run for 27 min at 120 V. The extracted band was purified using the QIAquick gel extraction kit (cat # 28704) as per manufacturer's protocol. 3) Annealing each pair of synthesized oligos. The oligo pairs were either annealed and phosphorylated or simply annealed without phosphorylation. For phosphorylation and annealing of the oligos, the thermocycler was set as 37^{o}C for 30 min, 95^{o}C for 5 min and then ramp down to 25^{o}C at 5^{o}C/min. For simply annealing the oligos, the setting was 95^{o}C for 5 min and then ramp down to 25^{o}C at 5^{o}C/min. The annealed oligos were diluted (1:200) in sterile water. 4) Ligation reaction. The dephosphorylated cut plasmid was ligated with the phosphorylated oligos and the simply cut plasmid was ligated with the annealed oligos without phosphorylation. As a negative control we had water instead of the oligos with the digested vector.

Table 3.1. **Detail of the cloning procedure.**

Vector digestion and dephosphorylation 30 min at 37°C	
lentiCRISPRv2	5 ug
FastDigest BsmBI	3 ul
FastAP	3 ul
10X FastDigest buffer	6 ul
100 mM DTT freshly made	0.6 ul
Nuclease free water	Made upto 60 ul total reaction
Vector digestion for 1 hr at 55°C	
lentiCRISPRv2	5 ug
NEB BsmBI	5 ul
10X NEB buffer	6 ul
Nuclease free water	Made upto 60 ul total reaction
Oligo pair annealing and phosphorylation	
Oligo 1 100 uM	1 ul
Oligo 2 100 uM	1 ul
10X T4 ligation buffer NEB	1 ul
T4 PNK NEB	0.5 ul
Nuclease free water	Made upto 10 ul total reaction
Oligo pair annealing	
Oligo 1 100 uM	1 ul
Oligo 2 100 uM	1 ul
Nuclease free water	Made upto 10 ul total reaction
Ligation reaction at RT for 10 min	
BsmBI digested plasmid dephos/ no dephos	50 ng
Diluted oligo duplex phos/ no phos	1 ul
2X Quick ligase buffer NEB	5 ul
Nuclease free water	Made upto 10 ul total reaction
Quick Ligase NEB	1 ul

5) Transformation into Stbl3 bacteria. 25 ul Stbl3 cells were mixed with 5 ul ligation reaction by gentle flicking. Cells were incubated on ice for 15-20 min. For transformation, cells were given heat shock at 42^{o}C for 30-40 s and then immediately put on ice for 2 min. 200 ul SOC media was added to the cells and shaken at 250 rpm for 1 hour. All the cells were plated on agar plates with appropriate selection reagent and incubated at 37^{o}C overnight for colony formation. After 16 to 18 hours, if colonies appeared, 3 to 4 random colonies were picked for 5 ml overnight cultures. The agar plates were then stored at 4^{o}C. Plasmids were extracted from the overnight cultures and colony PCR was performed with appropriate oligo 1 as forward and Cas9 as reverse primer with 2 ul cells. The region of the plasmid containing the cloned sgRNAs was PCR amplified and the sequence checked using Sanger sequencing.

Transient transfection into 293T cells

On Day 1, 5X10^{5} cells/well were plated in a 6-well plate. On day 2 transfection mix was prepared containing the plasmid (1.5 ug/well) and fugene (transfection reagent). In the transfection mix the ratio of fugene to plasmid DNA was 6:1. 100 ul of the transfection mix was dispensed in each well. On Day 3 media with the selection marker (puromycin 1 ug/ml) was added to the wells. Cells were selected for 72 hours and then used for the experiments.

T7 endonuclease I assay

Genomic DNA (gDNA) was extracted from the cells using the Blood and Cell culture DNA midi kit (Qiagen #13343) and the sgRNA target locus was PCR amplified (Table 3.2). The PCR products were run on 1% agarose gel to verify the size and amplification. PCR products were purified, and their concentrations determined. These PCR products were then set for the annealing reaction. 1 ul of T7 endonuclease I enzyme (NEB cat # M0302) was added to 19 ul annealed PCR products and incubated for 15 min at 37^{o}C. The reaction was stopped with 1.5 ul of 0.25 M EDTA. One may also consult the protocol provided in the NEB website. The fragmented PCR products were

analyzed on 10% TBE gel run in 0.5X TBE buffer for 70 min. The gel was then stained in 75 ml 0.5X TBE buffer with 0.7 ul EtBr for 20 min.

Preparation of stable MCF 10A p53 mutant cell lines expressing sgRNAs to knock-out PTEN

Once the sgRNAs targeting *PTEN* were cloned into lentiCRISPR v2 plasmid and the sequence verified, the cloned plasmids were purified. The lentivirus was prepared and used for transduction of the different MCF 10A cell lines as described above. The transduced cells were selected for at least 10 to 14 days before using them for the experiments. Please see the method section on

Table 3.2. **Detail of PCR and annealing for T7 endonuclease assay**

PCR component	50 ul reaction	Step	Temperature	Time
Sapphire 2X master mix	25 ul	Initial denaturation	98°C	30 s
10 uM forward primer	2.5 ul	35 cycles	98°C	5 s
10 uM reverse primer	2.5 ul		50-72°C depending on the Tm of primers	10 s
Template DNA	100 ng in total		72°C	20 s
Nuclease free water	Upto 50 ul	Final extension	72°C	2 min
		Hold	4°C	

Component	19 ul annealing reaction	Step	Temperature	Ramp rate	Time
DNA	200 ng	Initial denaturation	95°C		5 min
10X NEBuffer 2	2 ul	annealing	95-85°C	-2°C/sec	
Nuclease free water	Upto 19 ul		85-25°C	-0.1°C/sec	10 s
		Hold	4°C		

'*Generation of PTEN null MCF 10A cell-lines using shRNA'* for the detail of lentivirus generation and transduction of cells with the lentivirus.

Extraction of invasive cells

For the extraction of invasive cells, the 6-well BioCoat Matrigel invasion chamber plates (cat # 354481) were used. 1 million cells in 2 ml were seeded in each Matrigel coated insert in serum free media, making a total of 6 million cells per plate. In the bottom well, 2.5 ml media with serum was placed. After seeding, the plates were incubated at 37^{o}C and 5% CO_2 for 24 hrs. Following incubation, media from both the insert and the well were removed. The inserts and the wells were washed with 1X PBS. Using a cotton swab the cells on top of the insert, the non-invasive cells, were scraped off. 1 ml of trypsin was added to the bottom well and the insert now with only the invasive cells was placed over it. After 15 to 20 min (if the typsinization was complete, cells would be completely round and floating) media containing serum was added to the well to neutralize trypsin. The invasive cells were collected from each well in a tube and then re-plated in a new sterile 6-well plate. Images were taken at each step of the process. Once the invasive cells started to grow, they were transferred to petri-dishes for expansion. Pellets were made for genomic DNA (gDNA) extraction and vials for frozen cells were also prepared. Genomic DNA was extracted using the Blood and Cell culture DNA midi kit from Qiagen (cat # 13343) as per the manufacturer's protocol.

Sequencing of the PTEN locus

For sequencing of the targeted locus of *PTEN*, the locus was PCR amplified with suitable primers as per the Table 3.3, purified and submitted for MiSeq 2x150 bp run. To verify the disruption of the targeted gene, paired-ended sequences were aligned to the reference human genome (GRCh38.p92/hg38) using Bowtie2 (Bowtie2 2.1.0) (Langmead B, 2012). A deletion was called if

a deletion site was within ± 5bp apart from the sgRNA targeting site. Igvtools (igv_2.4.13) was used for data visualization (Thorvaldsdóttir 2013).

Soft agar assay

For the soft agar assay, the agar bed of 0.6% was made by mixing equal volumes of 2X media and 1.2% noble agar (autoclaved) and plating 800 ul for the base. The agar was then allowed to solidify. 10,000 cells/well in a 24 well plate was suspended in the soft agar with a final concentration of 0.3% noble agar. Each cell line was first trypsinized and then counted. To have 10,000 cells/well with final concentration of 0.3% agar, equal volumes of 20,000 cells in 2X media

Table 3.3. **Primers to amplify PTEN locus and the PCR detail.**

Set		Primer	
1	Forward Reverse	CTA CCC CTG TGC AGT TGA AAA TTC ACA TG CTG AAG TCC ATT AGG TAC GGT AAG CCA	Exon 2
2	Forward Reverse	CTG TTA AGT TTG TAT GCA ACA TTT CTA AAG TTA CC TAC TTG TCA ATT ACA CCT CAA TAA AAC TGA AGG	Exon 5
3	Forward Reverse	CTA CGA CCC AGT TAC CAT AGC AAT TTA GTG GTT CAA ATG CTT CAG AAA TAT AGT CTC CTG C	Exon 6

PCR component	50 ul reaction	Step	Temperature	Time
Sapphire 2X master mix	25 ul	Initial denaturation	98°C	30 s
10 uM forward primer	2.5 ul	35 cycles	98°C	15 s
			50-72°C depending on the Tm of primers	15 s
10 uM reverse primer	2.5 ul		72°C	15 s
Template DNA	2 ng/ul in total	Final extension	72°C	5 min
Nuclease free water	Upto 50 ul	Hold	4°C	

were mixed with 0.6% soft agar (first melted and then strictly maintained at 42°C) and 1 ml of the solution was put over the solid agar bed. At each step the agar was allowed to solidify, and formation of bubbles was avoided. We plated the MDA-MB-231 cells as the positive control and had a well without cells as the background control. 500 ul 1X media was added on top of the soft agar and the media was replenished every 3 to 4 days to allow the cells to form colonies. Colonies in soft agar appeared in 3 to 4 weeks or longer depending upon the cell lines used. Once the colonies were visible, the cells were fixed with 500 ul of 10% acetic acid and 10% methanol in water for 15-20 min and imaged under the light microscope.

Immunofluorescence

This protocol was used for staining cells in a 96-well format and a shaker was used during antibody incubations to obtain uniform staining. The cells were washed with 1X PBS (100 ul, 2 times) and fixed with 75 ul of 4% paraformaldehyde (VWR #AA43368-9M) per well for 15 min at room temperature (RT). The cells were washed with PBS and 75 ul of blocking buffer containing 5% goat serum (Life Technologies #PCN5000) with 0.32% Triton X-100 in 1X PBS was added per well for 1 hour at RT. After a PBS wash, 50 ul/well of primary antibody (at recommended dilutions in Ab protocol from manufacturer) in antibody buffer with 1% BSA and 0.3% Triton X-100 in 1X PBS was added and incubated at RT shaking gently at 350 rpm in an orbital shaker for 1 to 2 hours for same day processing or overnight at 4°C while shaking. Cells were washed with 1X PBS and 50 ul/well of secondary antibody (at recommended dilution) in antibody buffer was added. The 96-well plate was incubated at RT, shaking at 350 rpm for 1 to 2 hours. After a final wash with 1X PBS plate was stored at 4°C with 100µl 0.05% Sodium Azide (NaN_3 Sigma #S8032) in 1X PBS.

Results

PTEN protein knock-down using the shRNA system increases the invasiveness of p53 mutant cells

Deletion mutation in *PTEN* frequently co-appears with *TP53* missense mutations in invasive breast carcinomas (Kim 2015). We hypothesized that deletion of *PTEN* in the weakly invasive MCF 10A p53 mutant cells could enhance their invasive potential and demonstrate cooperativity between mutant PTEN and mutant p53.

To check the effect of reduced PTEN protein levels we tested three different shRNAs (PTEN sh1, 2, and 3) to knock-down the PTEN protein in the MCF 10A p53 mutant cells G245S, R273H and Y234C, which were the least invasive mutants as determined before using the transwell invasion assay (chapter 2) (Fig 3.5A). The MCF 10A cells overexpressing wildtype p53 protein (WT-OE) and the cells lacking p53 protein, the p53 null cells, were included as the controls for the experiment. The shRNA bearing plasmids were packaged into lentivirus, and the MCF 10A cell-lines were transduced with these lentiviruses. The cells were selected with puromycin for 7 to 10 days and lysates were prepared for western blots. All three cell lines expressing any of the shRNAs targeting *PTEN* showed decreased level of PTEN protein compared to cells without *PTEN* shRNA or cells expressing scrambled shRNA. Notably, the p53 null cells were produced by expressing p53-specific shRNA from the same vector as shRNA for PTEN. Thus, we could not select the PTEN shRNA-transduced cells with puromycin and no effective down-regulation of the PTEN protein was seen in the MCF 10A p53 null cells. We chose shRNA #3 for PTEN knock-down to determine any changes in the invasiveness of the cells because it shut down the PTEN protein levels more effectively than the rest of the shRNAs.

With PTEN protein knock-down there was an increase in the invasive potential of the weakly invasive p53 mutants G245S, R273H and Y234C (Fig 3.5B). The increase in invasiveness was

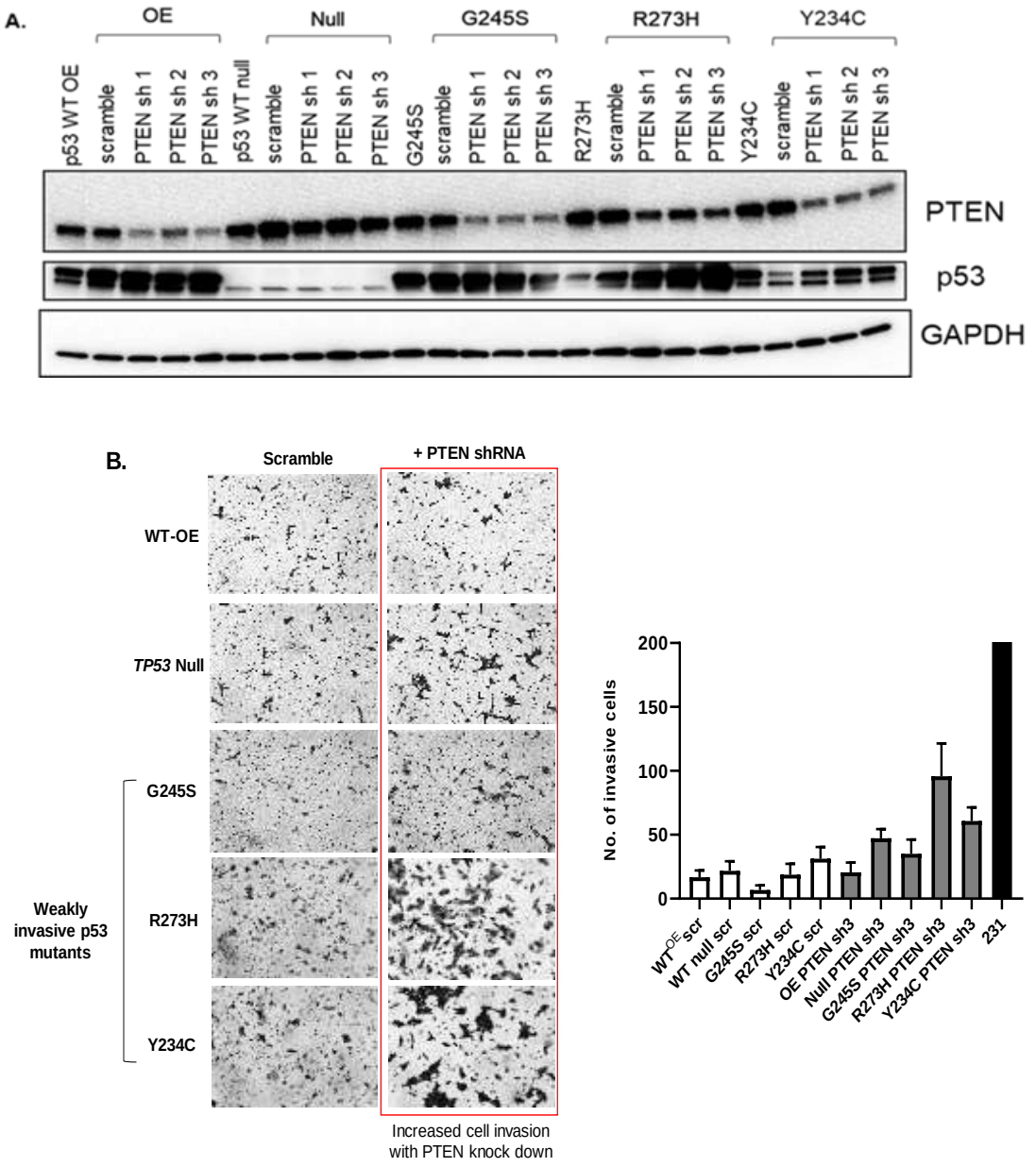


Figure 3.5. **Shutting down the PTEN protein using shRNAs**. A. Western blot showing the level of PTEN protein in the cells. GAPDH is the loading control. B. Transwell cell invasion assay results. PTEN knock-down increased the invasive potential of the weakly invasive MCF 10A p53 missense mutant protein expressing cell lines.

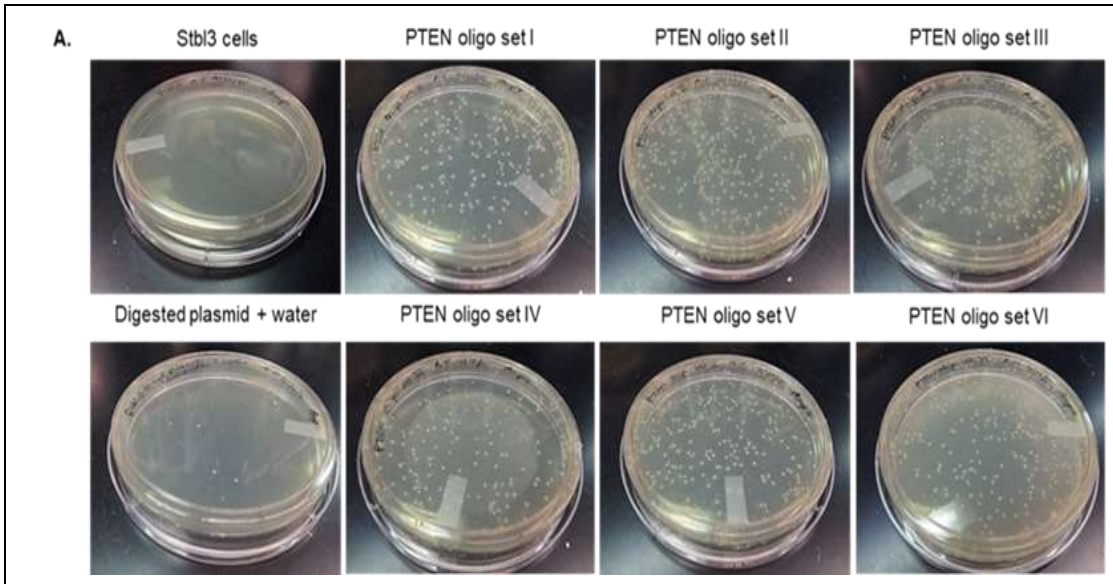
significantly higher for the R273H cells compared to G245S and Y234C cells. The MDA-MB-231 cells were included as the positive control, which is a highly invasive and metastatic TNBC cell

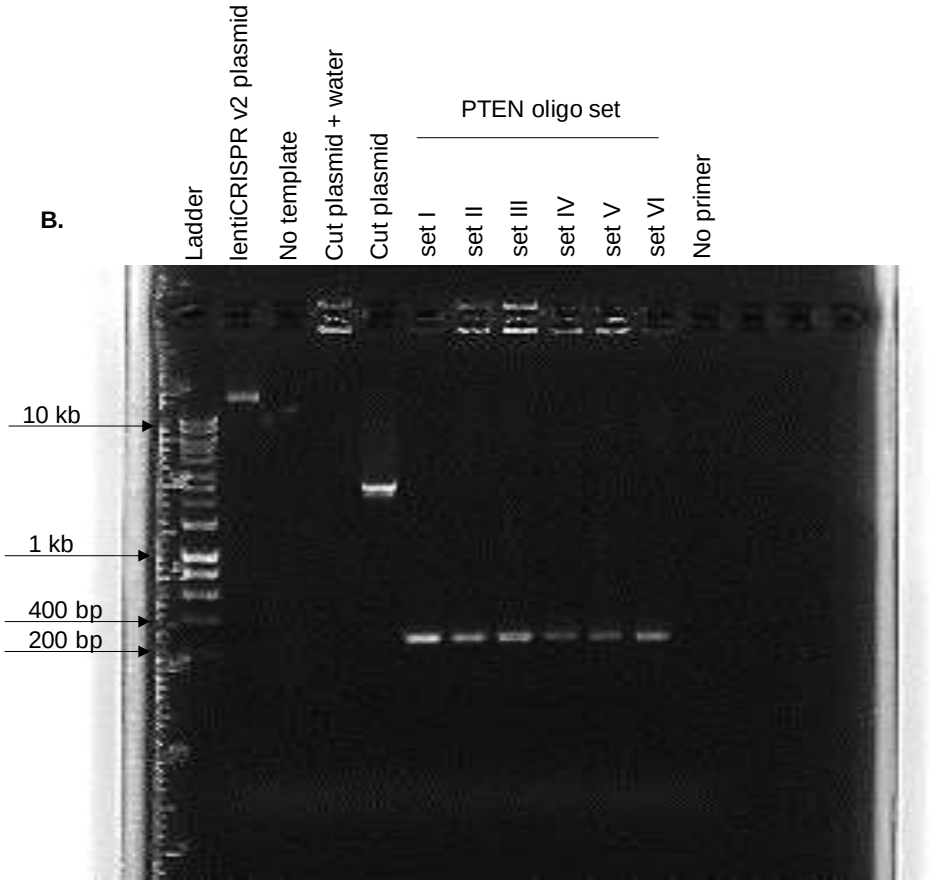
line. The results from the invasion assay demonstrated that PTEN loss enhances the invasion of p53 mutant cells across the Matrigel barrier, making the loss-of-function mutation of *PTEN* a 'co-driver' (co-existing driver mutation) of mutant p53.

Generation of PTEN knock-out using the CRISPR-Cas9 system

In the basal-like breast cancer (BLBC) or TNBC subtypes (chapter 1, page 29), *PTEN* deep deletions co-occur (5%) with *TP53* mutations. We used the CRISPR-Cas9 gene editing system to knock-out the *PTEN* gene and mimic the somatic mutations present in tumors. We designed a total of 6 sgRNAs (short guide RNAs) targeting different regions of the *PTEN* gene (Appendix A). These sgRNAs were chosen based on the parameters recommended by the online CRISPR design tools developed by the Broad Institute. We also verified that the sgRNAs did not target *PTENP1*, the pseudogene of *PTEN*, with similar sequences. These sgRNAs were first cloned into the CRISPR-Cas9 based lentivector lentiCRISPR v2, and the sequence were verified by Sanger sequencing before using them for further experiments (Fig 3.6).

To check the efficiency of the sgRNAs to knock-out *PTEN*, we transiently transfected the 293T cells with the sgRNA plasmids. The cells were harvested at 24, 48 and 72 hrs post transfection for lysate preparation and genomic DNA extraction to confirm the knock-out at the protein and genome levels, respectively. A reduction in the PTEN protein level was observed at 72 hrs post transfection with the PTEN sgRNA 4 (Fig 3.7A). The *PTEN* locus was amplified from the genomic DNA (gDNA) and digested with T7 endonuclease I to detect the genome editing efficiency by the sgRNAs. T7 endonuclease I detects and cleaves mismatched DNA. Amplification of the targeted locus followed by denaturation and slow annealing of the strands cause bulges resulting from mismatched DNA due to gene editing by the CRISPR-Cas9 system. These mismatches are recognized and then cleaved by T7 endonuclease I forming multiple DNA fragments. Multiple DNA fragments were seen on the 10% TBE gel from the 293T cells transfected with sgRNAs and none for the non-transfected 293T cells (Fig 3.7B).


A.
Stbl3 cells
PTEN oligo set I
PTEN oligo set II
PTEN oligo set III
Digested plasmid + water
PTEN oligo set IV
PTEN oligo set V
PTEN oligo set VI


B.
Ladder
lentiCRISPR v2 plasmid
No template
Cut plasmid + water
Cut plasmid
PTEN oligo set
set I
set II
set III
set IV
set V
set VI
No primer
10 kb
1 kb
400 bp
200 bp

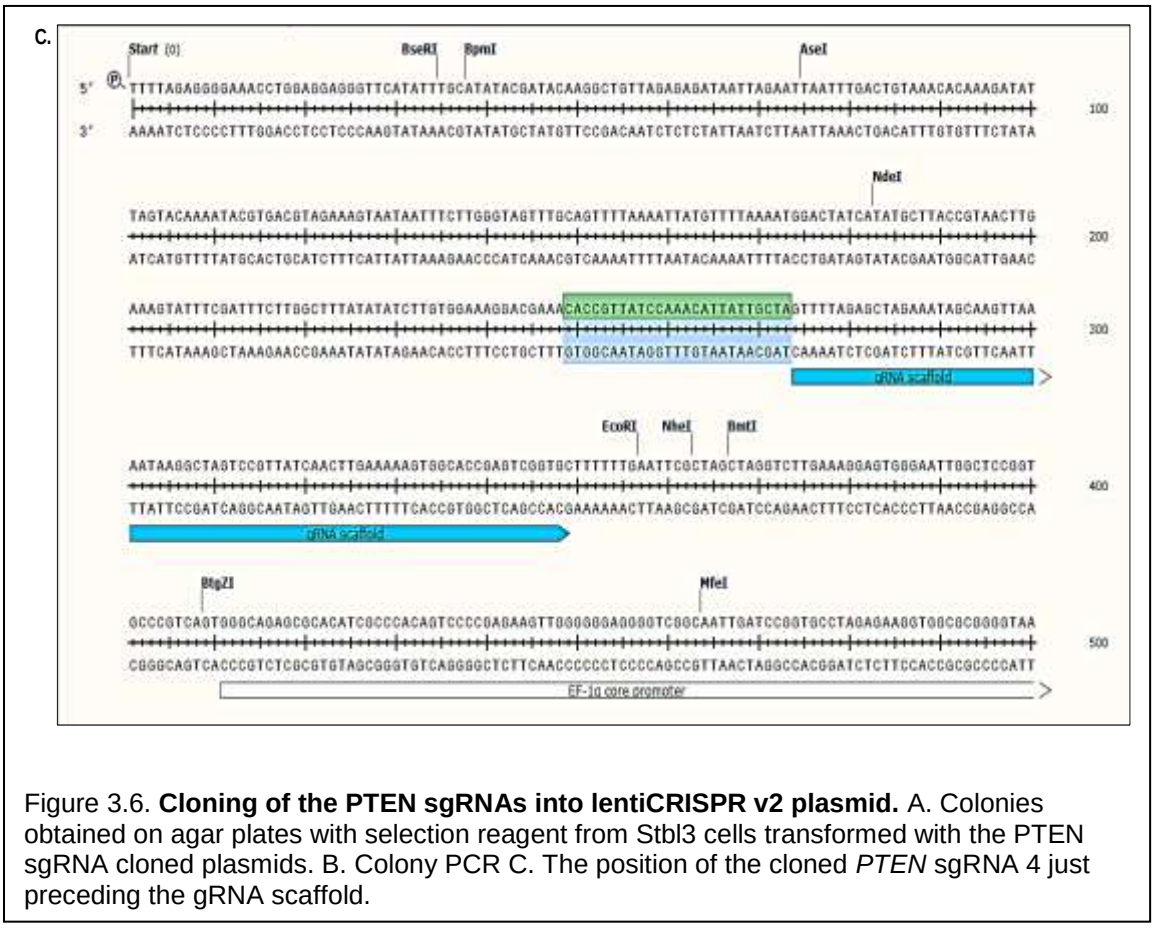


Figure 3.6. **Cloning of the PTEN sgRNAs into lentiCRISPR v2 plasmid.** A. Colonies obtained on agar plates with selection reagent from Stbl3 cells transformed with the PTEN sgRNA cloned plasmids. B. Colony PCR C. The position of the cloned *PTEN* sgRNA 4 just preceding the gRNA scaffold.

To transduce the MCF 10A p53 mutant cells, we prepared lentivirus expressing sgRNA 4, sgRNA 5 and sgRNA 6 cloned plasmids. We chose these 3 sgRNAs because of the importance of their targeting positions on the *PTEN* gene (Fig 3.8A). The sgRNA 4 targeted *PTEN* exon 2 and sgRNA 5 targeted exon 5. Both sgRNAs targeted the phosphatase domain of *PTEN* while sgRNA 6 targeted the linker region that connects the phosphatase domain with the C2 domain. The C2 domain positions the protein on the plasma membrane, and both the domains are essential for the activity of the PTEN protein. The MCF 10A cells expressing WT p53, WT-OE, G245S, R273H and Y234C p53 mutant cells were transduced with the PTEN sgRNA (4, 5 and 6) containing lentiviruses. The cells were selected for 2 weeks and then used for the experiments.

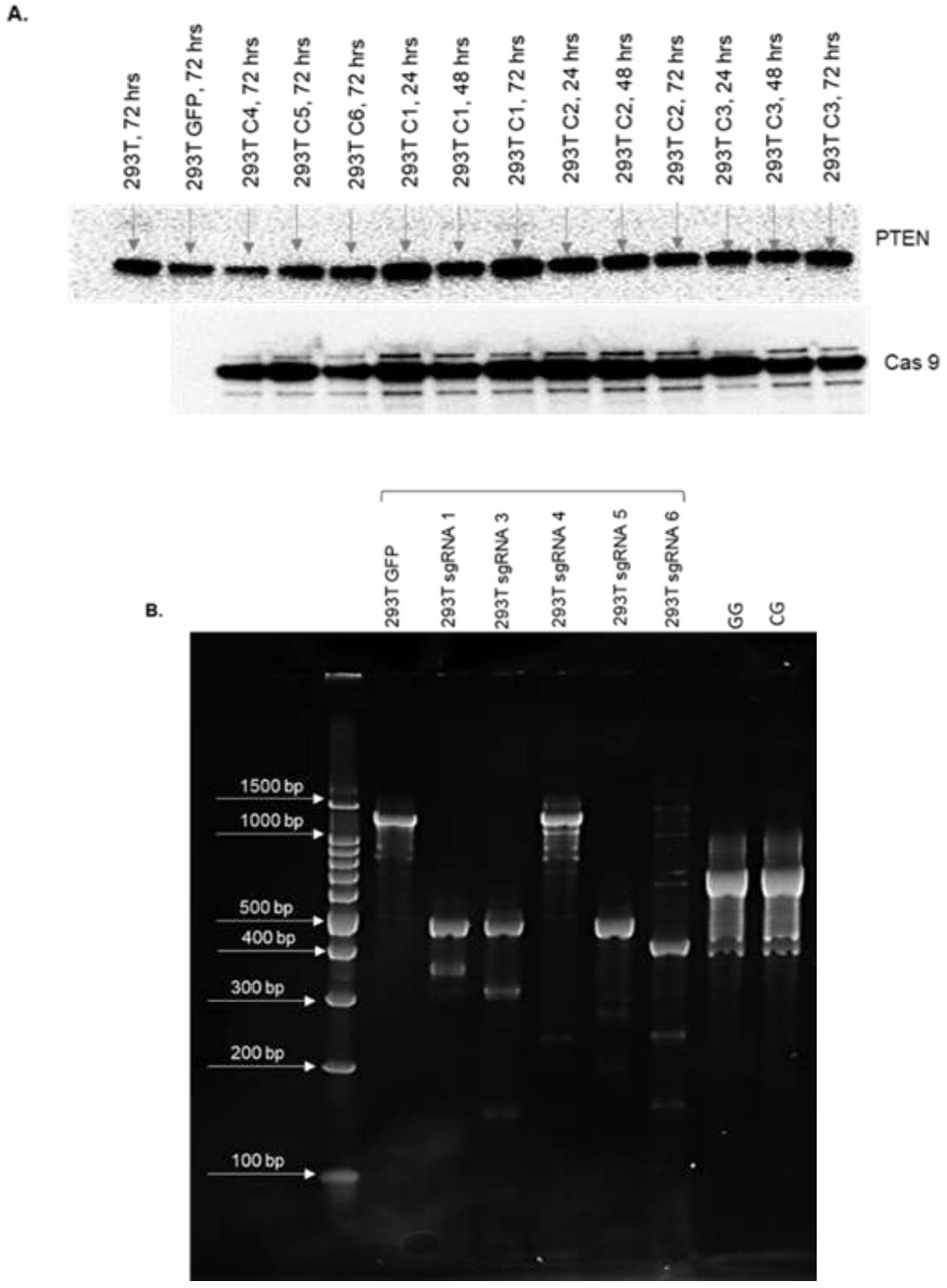


Figure 3.7. **Results from the transient transfection of 293T cells with PTEN sgRNA cloned plasmids.** A. Western blot to show PTEN protein levels. B. DNA fragments detected from T7 endonuclease I assay.

When the cell lysates from the culture were prepared to check the PTEN protein levels (Fig 3.8B), no reduction in the PTEN protein level was observed for any of the PTEN sgRNA-transduced cells which can be explained by the low efficiency of the CRISPR editing ranging between 20 and 30%. As the knock-out could have happened in a small fraction of the cell population, we did not observe reduction of PTEN protein by analyzing the bulk population of cells. Therefore, we used the invasion assay platform to enrich the clonal population of cells with *PTEN* knock-out by taking advantage of their increased invasiveness, as shown by the PTEN shRNA knock-down (Fig 3.5).

Method for enrichment of post-invasion cell population and confirmation of targeted gene deletion: foundation for the genome wide co-driver screen

The PTEN sgRNA-transduced Y234C cells were seeded in the matrigel coated Boyden chambers of the transwell invasion plates (Fig 3.9A). After the incubation period, the non-invasive cells were first scraped off from the top of Matrigel coated insert, and the invasive cells were detached with trypsin treatment. Only about 10% of cells invaded the Matrigel coated membrane. To yield enough nucleic acid from cells for sequencing of the *PTEN* locus to detect genome editing, the extracted invasive cells were expanded in culture for 4-5 doublings. We noted that the invasive cells had completely different morphology (Fig 3.9B) from the population of the pre-invasive Y234C PTEN sgRNA 4, 5 and 6 (denoted as PTEN C4, C5 and C6) cells.

Post-invasion cells were more elongated and spindle-like, characteristic of mesenchymal cells. The change in morphology was most pronounced in Y234C PTEN C4 cells. In the western blot, there was complete absence of the PTEN protein in the post-invasion Y234C PTEN C4 cells (Fig 3.9C), demonstrating that sgRNA 4 was effective in disrupting the PTEN gene, and we could enrich the true PTEN knock-out clones from the mixed population of cells using the invasion assay. We sequenced the *PTEN* locus from the genomic DNA isolated from the Y234C PTEN C4, C5 and C6 cells. Sequencing results confirmed a uniform deletion of 2 or 4 bases (*i.e.*, frameshift deletions) at the locus targeted by sgRNA 4. The targeted locus was in the exon 2

(close to the N-term) of the PTEN protein, and the efficient editing by sgRNA 4 resulted in frameshift mutation terminating the PTEN protein production (Fig 3.9D). On the contrary, sgRNA 5 and 6 failed to knock-out the PTEN gene.

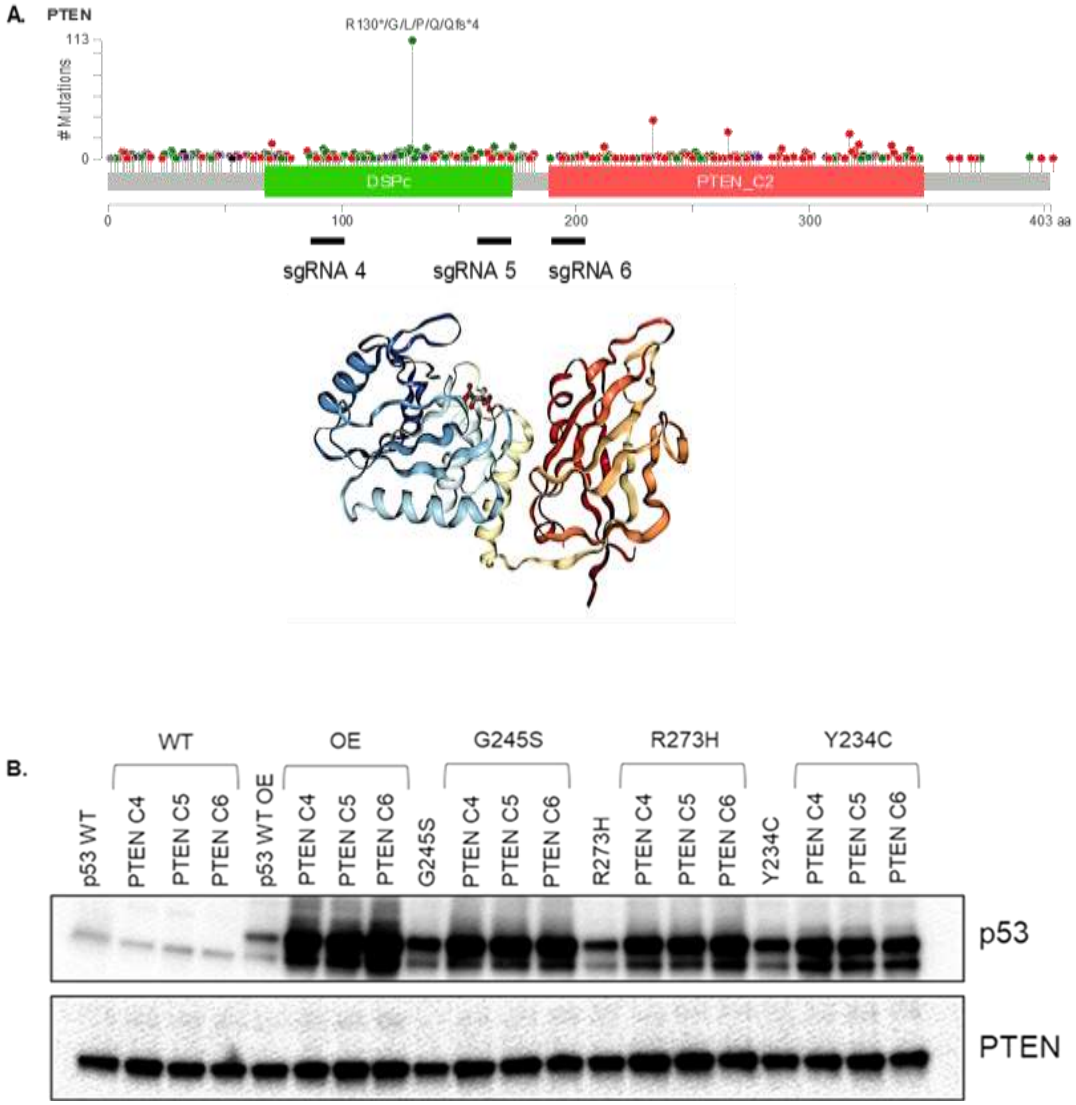
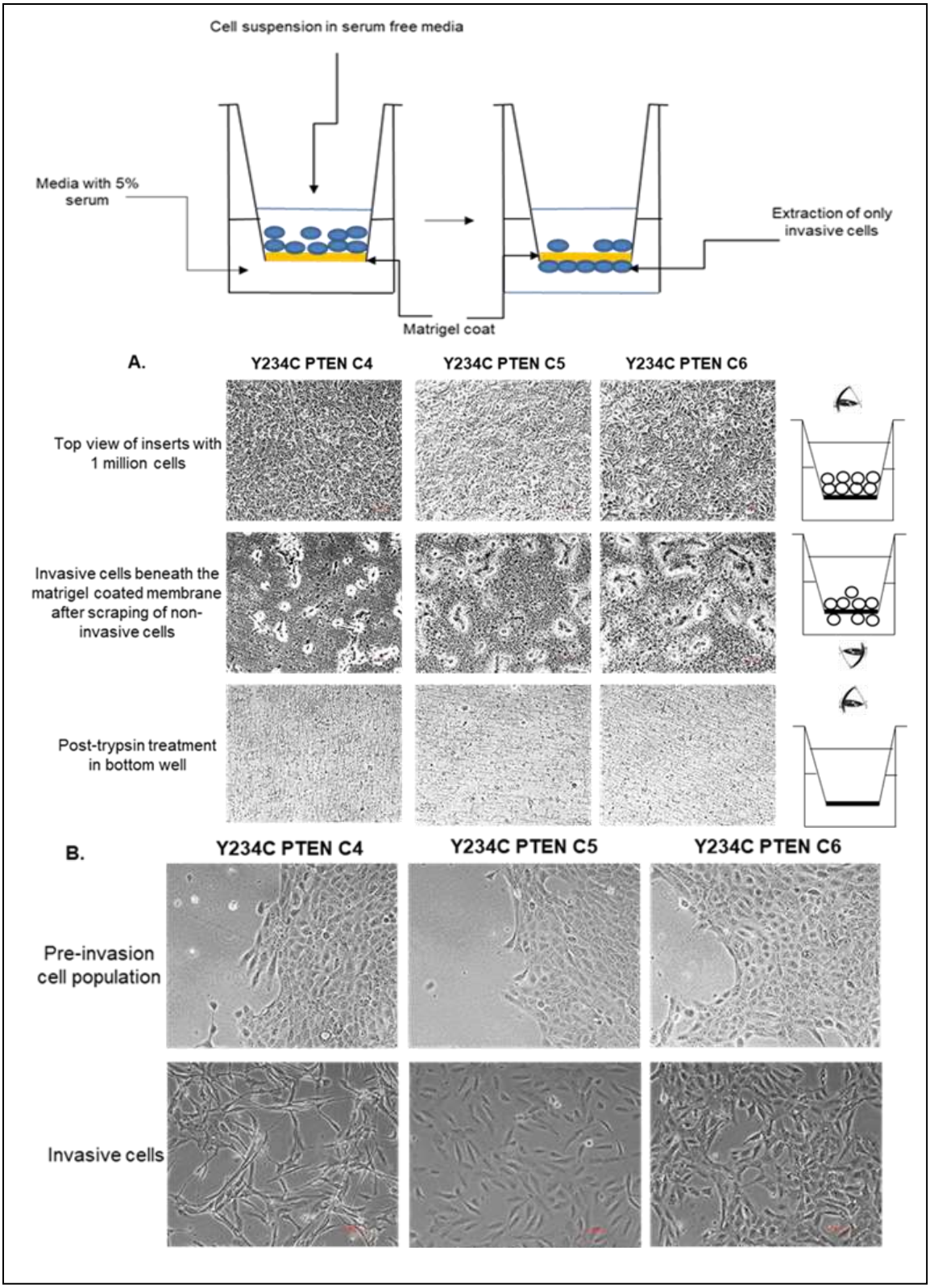


Figure 3.8. **sgRNA targeting positions on the *PTEN* gene.** A. sgRNA 4 and sgRNA 5 targeted the PTEN phosphatase domain and sgRNA 6 targeted the linker region between the phosphatase and the C2 domain. Structure of PTEN from PDB. B. Western blot to check PTEN protein levels. The cell lysates were prepared after transduction of cells with PTEN sgRNA expressing lentivirus. The transduced population was a mixed population where a very small fraction of cells lost PTEN due to CRISPR editing. Thus, in the bulk population of cells there was no reduction in PTEN protein level.


Cell suspension in serum free media
Media with 5% serum
Extraction of only invasive cells
Matrigel coat
A.
Y234C PTEN C4
Y234C PTEN C5
Y234C PTEN C6
Top view of inserts with 1 million cells
Invasive cells beneath the matrigel coated membrane after scraping of non-invasive cells
Post-trypsin treatment in bottom well
B.
Y234C PTEN C4
Y234C PTEN C5
Y234C PTEN C6
Pre-invasion cell population
Invasive cells

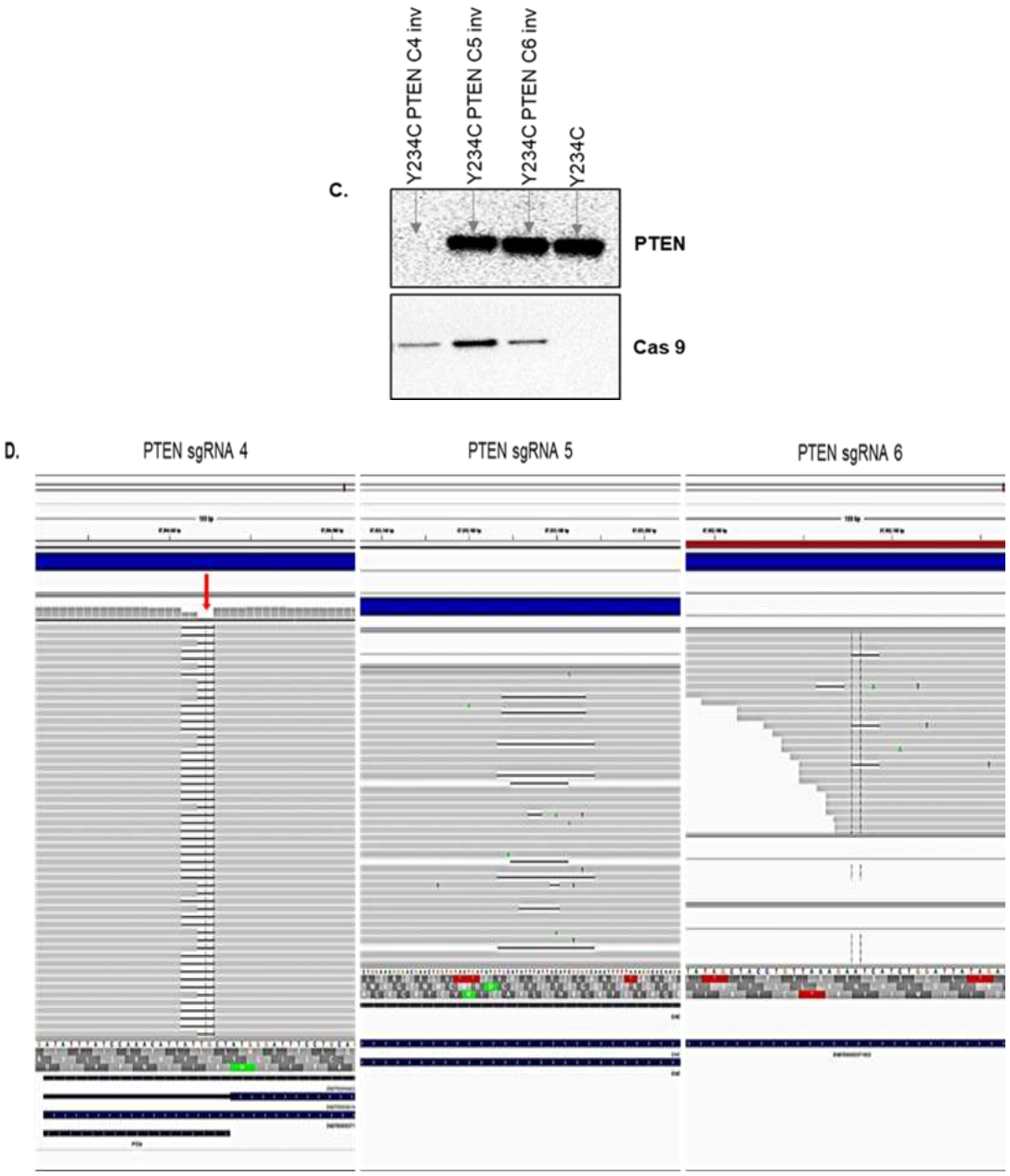


Figure 3.9. **PTEN gene knock-out by CRISPR-Cas9 editing.** A. Extraction of only the invasive cells that managed to breach the Matrigel coated inserts. C4, C5, C6 correspond to PTEN sgRNA 4, sgRNA 5 and sgRNA 6. B. Expansion and morphology of the invasive cells in 6-well plate. C. Western blot to check the PTEN protein levels. sgRNA 4 effectively knock-out PTEN gene. D. Sequencing results of the PTEN locus. The red arrow shows the position of the deleted base pairs. Uniform deletion of 2 to 4 bp resulted in frameshift mutation and subsequent deletion of the *PTEN* gene and the protein with sgRNA 4.

PTEN sgRNA 5 caused a non-uniform deletion of 2 to 22 bases at the exon 5, and sgRNA 6 resulted in non-uniform deletion of 4 bases at exon 6. This explained the presence of PTEN protein band and indicated a mixed population of Y234C PTEN C5 and C6 cells post-invasion. Ideally the cells expressing PTEN should not be present in the invasive fraction of cells. But it could be possible that such cells invaded along with the *PTEN* knock-out cells (Yamaguchi 2005, Weaver 2006), or there could be non-specific targeting by sgRNA 5 and 6 in the genome that directly or indirectly promoted the invasion of the cells. Off-target effects have been observed while using the CRISPR-Cas9 system for gene editing.

Immunofluorescent staining of the parental Y234C cells and the post-invasion Y234C PTEN C4, C5 and C6 cells with beta-catenin and vimentin showed different localization patterns of the proteins (Fig 3.10A). Beta-catenin is part of the cell adherens junction and is normally located at the cell-cell junction of the epithelial cells. For Y234C cells and the Y234C PTEN C6-inv (post-invasion or invasive Y234C cells transduced with PTEN gRNA 6) cells beta-catenin was located at cell-cell junctions (typical for epithelial cells), while the protein was delocalized from the cell-cell junctions and present in the cytoplasm in the Y234C PTEN C4-inv and C5-inv cells as typically seen in migratory or invasive cells. Vimentin, a mesenchymal marker more abundant in invasive cells, was observed throughout the Y234C PTEN C4-inv cells. Overall, the morphology, beta-catenin and vimentin staining patterns indicated that *PTEN* knockout in Y234C cells led to more mesenchymal phenotype and invasive behavior, which was especially apparent in the Y234C PTEN C4-inv cells.

The post-invasion Y234C PTEN C4 cells were also capable of anchorage independent growth and formed colonies in the soft agar (Fig 3.10B). While the post-invasion Y234C PTEN C5 cells failed to form colonies, a fraction of the post-invasion PTEN C6 cells formed colonies, which possibly had *PTEN* knocked-out. From the results, we chose the most effective sgRNA 4 to knock-out *PTEN* in the other p53 mutant cell lines.

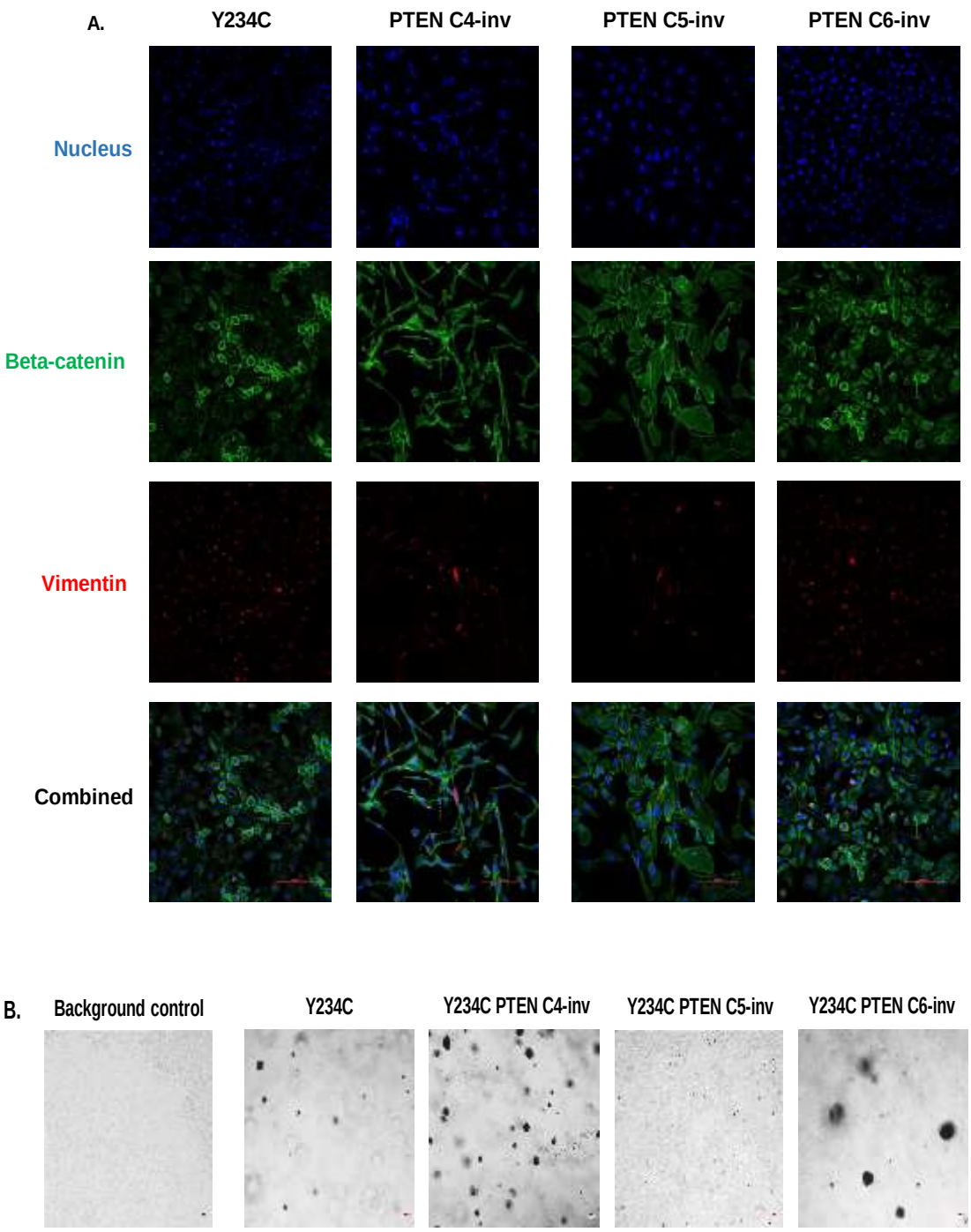


Figure 3.10. **Results from immunofluorescence and soft agar assay.** A. Immunofluoresence study showing localization of beta-catenin and vimentin in the post-invasion or invasive (inv) cells. Beta-catenin was in the cytoplasm than cell-cell junction for PTEN C4-inv cells indicating mesenchymal phenotype. B. Anchorage independent growth of post-invasion Y234C PTEN cells.

PTEN knock-out increases invasiveness of p53 mutant cells

Two other weakly invasive G245S and R273H cells were first transduced with PTEN sgRNA 4 lentivirus and then selected with puromycin for 2 weeks. As performed above for Y234C cells, the invasive population of cells were enriched by using the invasion chambers and confirmed for the reduced level of PTEN proteins (Fig. 3.11B). When we tested and compared the panel of cell lines transduced with PTEN sgRNA C4, the post-invasion or invasive MCF 10A p53 WT, WT-OE, G245S, R273H and Y234C cells were capable of colony formation in soft agar (Fig 3.11A).

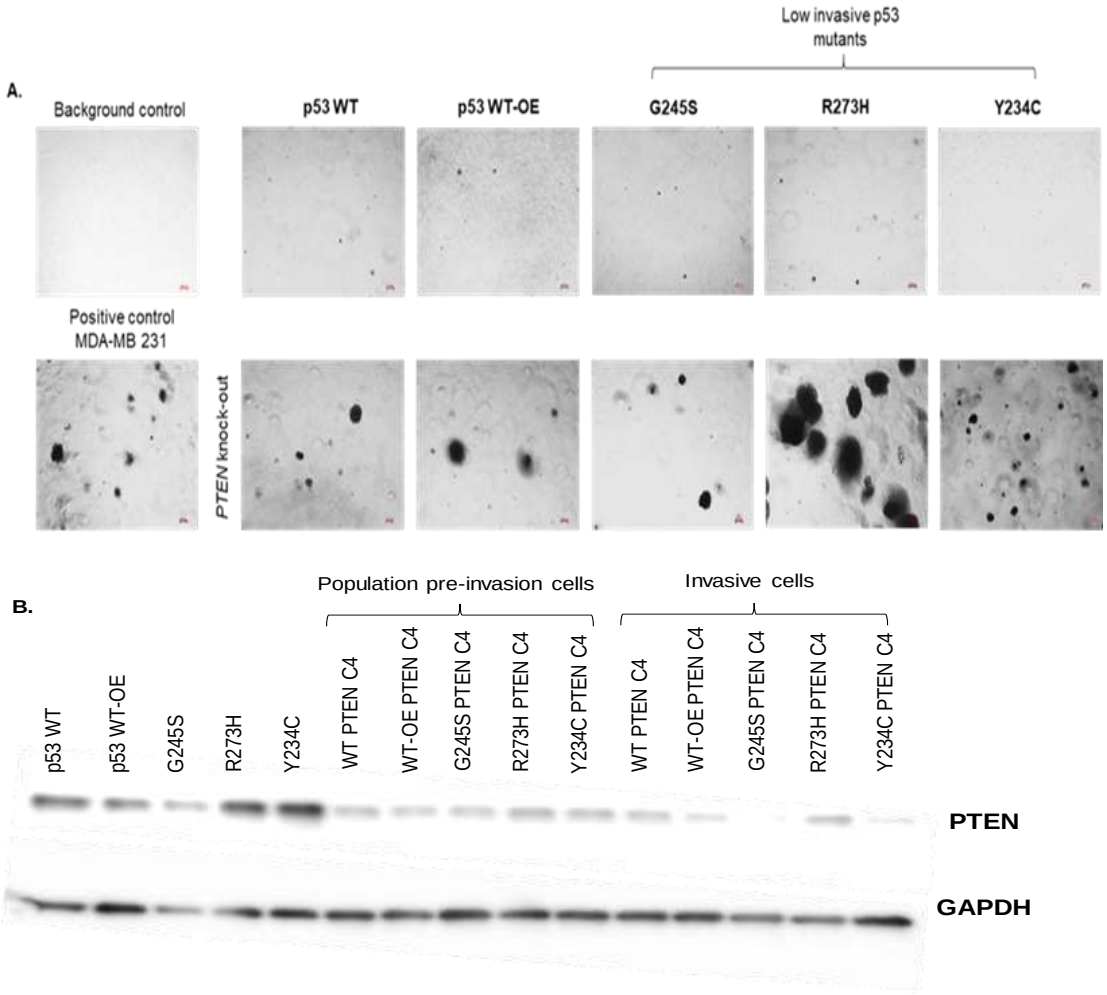


Figure 3.11. **Soft agar assay to show transformation of cells.** A. Colony formation in soft agar by cells expressing mutant p53 and knocked-out *PTEN*. B. Western blot to check the PTEN protein level in p53 mutant cells.

Like the results obtained with PTEN knock-down by shRNA (Fig 3.5), the effect of PTEN loss in R273H cells was most prominent among all the mutants. These cells were most invasive in the Boyden chamber/ transwell assay and formed huge spheres in the soft agar compared to G245S and Y234C mutants.

Conclusion

Loss of function of the tumor suppressor *PTEN* is frequently accompanied by the loss of function of *TP53* gene in the most aggressive subtype of basal-like breast cancer that lacks the detectable expression of ER, PR and HER2 receptors. The concurrent downregulation of these two tumor suppressor genes leads to increased cell invasion and anchorage independent growth *in vitro* and tumor formation *in vivo*. We sought to find out the effect of *PTEN* loss in the context of different *TP53* missense mutations, and the results recapitulated the previous findings but provided additional knowledge on the phenotypic heterogeneity among the cells with different p53 mutants and PTEN deletion.

We intended to setup a function-based approach to identify the distinct co-drivers of mutant p53 and required a positive control to set the conditions for the genome wide screen for the co-drivers. Deletion in *PTEN* is known to cooperate with mutant p53 in exacerbating cancer-related phenotypes such as cell proliferation and invasion. Hence, we chose to delete *PTEN* by CRISPR-Cas9 method in the MCF 10A mutant p53 expressing cells and observe the changes in cell invasion phenotype. This entire experimental setup was designed to establish a pipeline to identify co-drivers of mutant p53 by targeted gene deletions using a genome wide CRISPR based plasmid library and resulting change in cellular phenotypes. Using deletion in *PTEN* as a control, the known co-driver of mutant p53, we could perform phenotype-based enrichment of cells with targeted gene deletion, confirmation of gene deletion by NGS and validate the change in the observed phenotype of increased cell invasion.

PTEN protein knock-down using shRNAs in the weakly invasive MCF 10A mammary epithelial cells expressing p53 missense mutant proteins (G245S, R273H and Y234C) caused an increase in cell invasion compared to the cells with only p53 mutation. Particularly, invasion of the R273H cells was most influenced by the PTEN knock-down. Reduction in PTEN protein levels in MCF 10A cells with wildtype p53 also invaded more than the MCF 10A cells without PTEN knock-down.

As shRNA-based partial knock-down of PTEN proteins cannot recapitulate the complete loss of PTEN function via homozygous deletion observed in breast cancer patients, we utilized the powerful gene editing technology of CRISPR-Cas9 system. Among all the designed sgRNAs targeting different parts of the PTEN gene, the sgRNA targeting a region close to the N-terminus was found to be the most efficient in knocking-out *PTEN*. Targeting at this locus caused uniform deletion of 2 or 4 bps resulting in frameshift mutations and consequent ablation of the gene function. As the efficiency of gene editing by the CRISPR system is relatively low, we developed a phenotype selection-based strategy to enrich the cells with loss-of-function *PTEN* deletion, in which the invasive population of PTEN sgRNA-transduced cells were isolated after the invasion assay, followed by confirmation of gene deletion by the sequencing of the locus. The invasive cells expressing different p53 mutants with *PTEN* deletion presented a spindle-like morphology, characteristics of mesenchymal cells such as cytoplasmic β-catenin and vimentin staining, and capability of anchorage independent growth in soft agar. Consistent to the shRNA-based results, anchorage independent growth in R273H *PTEN* null cells was most prominent than the mutants G245S and Y234C.

These results collectively showed that simultaneous down-modulation of both *PTEN* and *TP53* synergize to promote aggressive cellular behavior but to different extents in the context of different *TP53* mutations. As mutant *TP53* is a known driver of breast cancer, and both *PTEN* and *TP53* mutations co-exist in basal type tumors, *PTEN* can be called a “co-driver” of mutant *TP53*.

In addition, the cell enrichment method that we developed for gene editing could be applied for screening for any given phenotype and to find the genotype that caused it.

In the next step with the defined parameters obtained from the PTEN deletion screen, we set up the genome wide screen for the discovery of co-drivers of different mutant p53 based on the cooperativity of the mutations in enhancing cell invasion.

References

Boettcher, Michael, and Michael T. McManus. "Choosing the right tool for the job: RNAi, TALEN, or CRISPR." *Molecular cell*58, no. 4 (2015): 575-585.

Canver, Matthew C., Daniel E. Bauer, Abhishek Dass, Yvette Y. Yien, Jacky Chung, Takeshi Masuda, Takahiro Maeda, Barry H. Paw, and Stuart H. Orkin. "Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells." *Journal of Biological Chemistry*289, no. 31 (2014): 21312-21324.

Chalhoub, Nader, and Suzanne J. Baker. "PTEN and the PI3-kinase pathway in cancer." *Annual Review of Pathological Mechanical Disease* 4 (2009): 127-150.

Georgescu, Maria-Magdalena, Kathrin H. Kirsch, Tsuyoshi Akagi, Tomoyuki Shishido, and Hidesaburo Hanafusa. "The tumor-suppressor activity of PTEN is regulated by its carboxyl-terminal region." *Proceedings of the National Academy of Sciences* 96, no. 18 (1999): 10182-10187.

Kim, Gwangil, Maria Ouzounova, Ahmed A. Quraishi, April Davis, Nader Tawakkol, Shawn G. Clouthier, Fayaz Malik et al. "SOCS3-mediated regulation of inflammatory cytokines in PTEN and p53 inactivated triple negative breast cancer model." *Oncogene* 34, no. 6 (2015): 671.

Lee, Jie-Oh, Haijuan Yang, Maria-Magdalena Georgescu, Antonio Di Cristofano, Tomohiko Maehama, Yigong Shi, Jack E. Dixon, Pier Pandolfi, and Nikola P. Pavletich. "Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association." *Cell* 99, no. 3 (1999): 323-334.

Lee, Yu-Ru, Ming Chen, and Pier Paolo Pandolfi. "The functions and regulation of the PTEN tumour suppressor: new modes and prospects." *Nature Reviews Molecular Cell Biology*(2018): 1.

Li, Andrew G., Landon G. Piluso, Xin Cai, Gang Wei, William R. Sellers, and Xuan Liu. "Mechanistic insights into maintenance of high p53 acetylation by PTEN." *Molecular cell*23, no. 4 (2006): 575-587.

Liu, Jeff C., Veronique Voisin, Sharon Wang, Dong-Yu Wang, Robert A. Jones, Alessandro Datti, David Uehling et al. "Combined deletion of Pten and p53 in mammary epithelium accelerates triple-negative breast cancer with dependency on eEF2K." *EMBO molecular medicine* 6, no. 12 (2014): 1542-1560.

Mayo, Lindsey D., and David B. Donner. "The PTEN, Mdm2, p53 tumor suppressor–oncoprotein network." *Trends in biochemical sciences* 27, no. 9 (2002): 462-467.

Pon, Julia R., and Marco A. Marra. "Driver and passenger mutations in cancer." *Annual Review of Pathology: Mechanisms of Disease* 10 (2015): 25-50.

Ran, F. Ann, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. "Genome engineering using the CRISPR-Cas9 system." *Nature protocols* 8, no. 11 (2013): 2281.

Rao, Donald D., John S. Vorhies, Neil Senzer, and John Nemunaitis. "siRNA vs. shRNA: similarities and differences." *Advanced drug delivery reviews* 61, no. 9 (2009): 746-759.

Sansal, Isabelle, and William R. Sellers. "The biology and clinical relevance of the PTEN tumor suppressor pathway." *Journal of clinical oncology* 22, no. 14 (2004): 2954-2963.

Shah, Sohrab P., Andrew Roth, Rodrigo Goya, Arusha Oloumi, Gavin Ha, Yongjun Zhao, Gulisa Turashvili et al. "The clonal and mutational evolution spectrum of primary triple-negative breast cancers." *Nature* 486, no. 7403 (2012): 395.

Stambolic, V., D. MacPherson, D. Sas, Y. Lin, B. Snow, Y. Jang, S. Benchimol, and T. W. Mak. "Regulation of PTEN transcription by p53." *Molecular cell* 8, no. 2 (2001): 317-325.

Weaver, Alissa M. "Invadopodia: specialized cell structures for cancer invasion." *Clinical & experimental metastasis* 23, no. 2 (2006): 97-105.

Yamaguchi, Hideki, Jeffrey Wyckoff, and John Condeelis. "Cell migration in tumors." *Current opinion in cell biology* 17, no. 5 (2005): 559-564.

CHAPTER 4

A FUNCTIONAL GENOMIC SCREEN IDENTIFIED DISTINCT INVASION PROMOTING 'CO-DRIVERS' OF DIFFERENT MISSENSE MUTANTS OF TP53

Introduction

As mentioned in chapter 3, there are two broadly used methods to differentiate the driver genes from the passengers. They are the frequency-based and the function-based approach. There are two major limitations associated with the frequency-based approach. This approach is not efficient in identifying the non-prevalent driver mutations, the co-drivers, as these mutations are not frequent or non-recurrent unlike the clonal drivers and hence, difficult to differentiate from the background mutation frequency. Due to the low frequency of the co-drivers in a given sample, a very large number of samples (tumors or cells) need to be sequenced to identify the possible co-drivers which is not cost effective. In contrast, the function-based approach to identify co-drivers is an efficient method as the co-drivers can be differentiated from the non-functional passengers based on a functional advantage making this approach more reliable than the frequency-based approach. Choosing the function-based method, we designed a genome wide screen to identify the possible 'co-drivers' of missense mutant p53. In the previous chapter (chapter 3) we described about a CRISPR based screening pipeline that could deliver gRNA/Cas9 into the MCF10A cells, isolate/enrich invasive cell populations, and validate the gene deletion by using *PTEN* as a control, which is a known co-driver of missense mutant p53 frequently observed in invasive breast carcinomas. This established pipeline can be used for a genome wide functional screen for the discovery of unknown possible co-drivers of missense mutant p53.

Aim

To identify the co-drivers of missense mutant p53 and to examine if each mutant p53 required a unique set of co-drivers to promote invasion. We designed a functional genomics screen to test the impact of deletion of every human gene on cell invasion in the cells expressing different mutant p53 proteins.

Approach

For the co-driver screens, we chose to utilize CRISPR-based approach, since shRNA-based method does not result in complete loss of function of the genes. CRISPR-based gene editing works at the level of DNA and directly recapitulates the effect of genetic alterations. The shRNAs degrade the mRNAs and work at the gene transcription level that eventually show reduced protein. In such a case, there could be partial deletion of protein that may render the results inconclusive. Gene knock-out by CRISPR-Cas9 system results in complete deletion of the protein. A genome wide CRISPR based screen that perturbs all the human protein coding genes to identify the possible co-drivers of mutant p53 promoting cell invasion is a more systemic and effective method than screening for selected individual genes. Though expression libraries can be used for such a purpose but in classical molecular biology, gene deletion is the method of choice to determine the function of a gene. Hence, we chose a gene knock-out library (commercially available at Addgene from Feng Zhang's lab) for screening of the co-drivers of mutant p53 that enhance cell invasion.

Genetic screens with CRISPR based pooled libraries

Genetic screens have become essential biological tools for unbiased functional genomics or phenotype screening, and the CRISPR-Cas9 system for genome editing is a very powerful tool

for the screening because of its easy programmability (i.e., easy to design sgRNAs for targeting genes) and flexibility (i.e., a variety of engineered Cas9 enzymes for many types of gene manipulation). The use of CRISPR system has diversified and is not just limited to gene knock outs. This technique has been adopted for transcriptional repression (CRISPRi) with 90-99% repression with limited off-target effects and transcriptional activation (CRISPRa) with endonuclease dead Cas9 with gene expression modulation of over 1000-fold range (Gilbert 2014). The other applications also include the use of CRISPR-Cas9 for genetic sequence replacement or insertion, histone modification, DNA methylation and imaging of specific genomic loci. Therefore, the pooled libraries of single guide RNAs (sgRNAs) have enabled a wide range of different types of genome wide screening for coding genes, gene regulatory elements and non-coding genome features in given biological contexts. Such pooled sgRNA libraries are easy and inexpensive to synthesize, can be cloned into pooled plasmid library, and used for virus production and then cell screening *in vitro* or *in vivo* (Shalem 2015).

The pooled libraries generally contain multiple sgRNAs targeting one gene and produce very reliable distinct phenotypes and have been used for both positive and negative selection of phenotypes. The CRISPR-Cas9 genetic screens have successfully identified and characterized genes conferring resistance to treatments (Shalem 2014), enhancer elements (Korkmaz 2016) and long non-coding RNAs (Zhu 2016) in the human genome. For the pooled screens, virus transduction needs to be at low multiplicity of infection (LMOI) and NGS is needed for the screen read-out. Transduction of cells at low MOI ensures that each cell receives a single sgRNA thus, manipulation of one gene. This allows NGS of pooled cells or cell populations. In a typical screen, comparison of results between a reference cell pool (control) and the experimental cell pool (with treatment) identifies the genotype (sgRNAs) resulting in the selected phenotype.

The animal models of cancer closely mimic the human disease and direct *in-vivo* CRISPR screens in the small animals (like mice) have identified the causative mutations driving cancer in the tissue microenvironment and understanding the clinical transformation of cancer. For

example, in a genome wide CRISPR screen in mouse model for lung tumor metastasis showed effect of distinct mutations on primary tumor growth, late stage primary tumors and lung metastases (Chen 2015). Such a method can also be used to test the gene-phenotype relationships during cancer evolution. Direct delivery of sgRNAs in adeno-associated virus (AAV) vector in the lungs caused disruption of *TP53* and *LKB1* and point mutations in *KRAS* (knock-in mutation) leading to adenocarcinomas (Platt 2014). CRISPR-mediated genome editing has unraveled the combinations of mutations that influence tumor growth pattern, tendency to metastasize and resistance to therapy (Chow 2018).

Genome-scale CRISPR knock-out (GeCKO) v2.0 pooled libraries

We chose a commercially available CRISPR-based library, the GeCKO v2.0, for the genome-wide screen of the co-drivers of different missense mutant p53. This GeCKO v2.0 library was developed in the Zhang lab (Shalem 2014) and commercially available at Addgene. The library contains type II CRISPR nuclease system for genome editing in mammalian cells in a pooled library targeting the early consecutive exons and comprises of a total of 122,417 unique sgRNAs targeting the entire human genes. Each plasmid in this library is the lentiCRISPRv2 backbone vector and contains a 20 bp gene-specific sgRNA (single gRNA) for gene knock-out. To reduce the library complexity, the GeCKO library is split into two half-libraries A and B, and each has 3 sgRNAs per gene, making a total of 6 sgRNAs per gene. Both the half-libraries also contain 1000 control sgRNAs without any sequence homology to human genome. Library A targets miRNAs with 4 sgRNAs per miRNA. The lentiCRISPRv2 vector produces a very high viral titer, and the sgRNAs were designed to minimize off-target effects, making this library ideal for phenotype screening.

Library coverage for efficient screening

Before the screening for the co-drivers, it was essential to ensure the full representation of the library during its amplification, lentivirus production and during the transduction of the cells. A critical parameter to consider in performing any genome wide screen is whether the experimental design such as the number of cells and virus particles will be enough to test evenly every gene and identify the hits. In this study, we targeted for a 250x coverage per sgRNA, i.e., one sgRNA was represented in an average of 250 cells and tested for invasion promoting activity during the genome wide screening process. As the transduction of cells was carried out at a low MOI (multiplicity of infection) of around 0.3, we calculated the number of cells required based on the estimated transduction efficiency of 20 to 40% (Table 4.1). We also estimated the number of viral particles (virus particles/ no. of cells equated to 0.3 (the chosen MOI)) needed for the low MOI transduction and scaled up the lentiviral library production process accordingly.

Table 4.1. **Calculation to maintain 250-fold library representation.**

Half library	Total sgRNA constructs (GeCKO v2 library)	No. of cells for 250 fold representation	Cells screened at MOI=0.3	Lentiviral particles needed
A	65,383	16,345,750	65,383,000	19,614,900
B	58,028	14,507,000	58,028,000	17,408,400

PCR amplification of genome integrated sgRNA cassette

In order to find out which gene deletions cause the phenotypic changes after phenotype selection we designed a PCR-based strategy by taking advantage of the fact that every sgRNA containing plasmid share common DNA sequences except the sgRNA sequences. Therefore, the genome-integrated sgRNA cassettes can be amplified with the universal primer pair and sequenced by next-generation sequencing (NGS), from which each sequence read can be mapped to each

sgRNA/gene and counted. This method was preferred over the whole-genome sequencing, since it is less time consuming and cost effective. The method simplifies the analysis of results from the screen. Comparison of sgRNAs in the reference (pre-invasion) cell pool and the experimental (post-invasion) yields the enriched sgRNAs corresponding to the genes whose deletion or loss of function resulted in increased cell invasion of the p53 mutant cells. The genomic region containing the integrated sgRNA was amplified from genomic DNA using a two-step PCR method to derive short amplicons (300 bp) that can be sequenced directly. A 700 bp region was first amplified using a primer set (primer set 2, PCR 1) with primers specific to the lentiCRISPR v2 plasmid and a defined PCR program. Using the purified PCR 1 product as the template, PCR 2 was performed to get a 300 bp amplicon (Fig 4.1). This 300 bp amplicon was sequenced, and the read counts for each sgRNA was normalized against the total number of reads for each sample to CPM (counts per million reads). Lastly, by selecting the sgRNAs with the largest ratio of CPM in post-selection population over the ones in pre-selection population, we could identify potential co-driver genes.

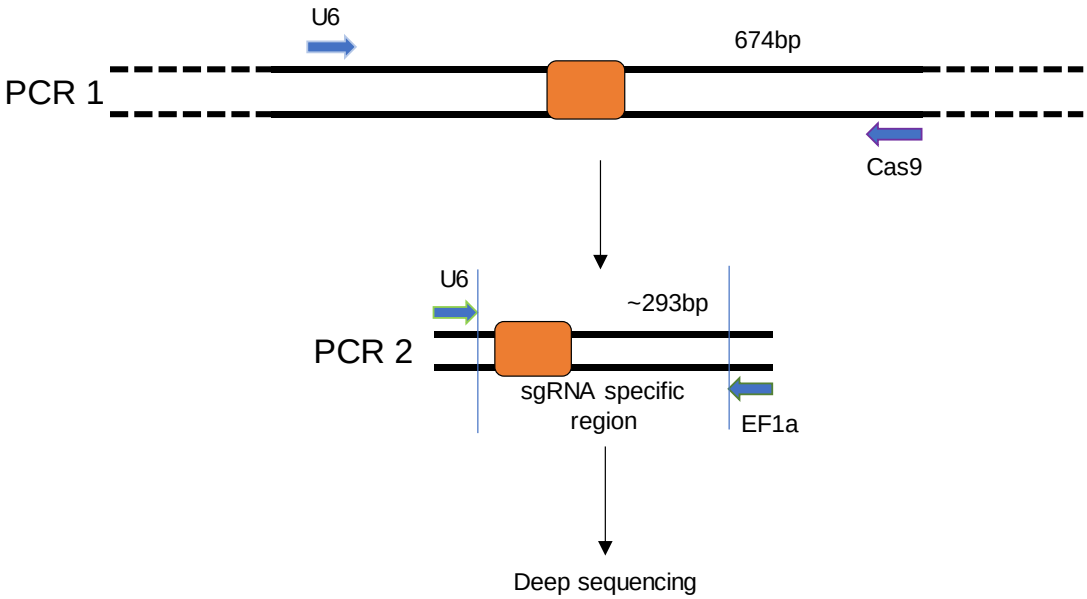


Figure 4.1. **PCR amplification to capture genome integrated sgRNA.** The region containing the sgRNA was amplified from genomic DNA by a 2 step PCR. The 300 bp amplicon was then sequenced.

The genome wide 'co-driver' screen design

Leveraging the optimized invasion assay condition in Chapter 3, we designed a screening protocol that allows enrichment of the cells with increased invasion after sgRNA library transduction (Fig 4.2). The library-transduced cells (*i.e.*, pre-selection or reference pool) were plated in the invasion chambers with a Matrigel-coated membrane and formed the pre-invasion reference cell pool. After incubation, the non-invasive cells were scraped off from inside the chambers. The invasive cells that crossed the Matrigel-coated membrane (*i.e.*, post-selection or invasive population) were isolated by adding trypsin in the bottom well. sgRNA cassettes were then PCR-amplified from each pool (reference and post-invasion) and sequenced, and the fold enrichment of sgRNAs in the invasive pool over the reference pool were calculated. The enriched

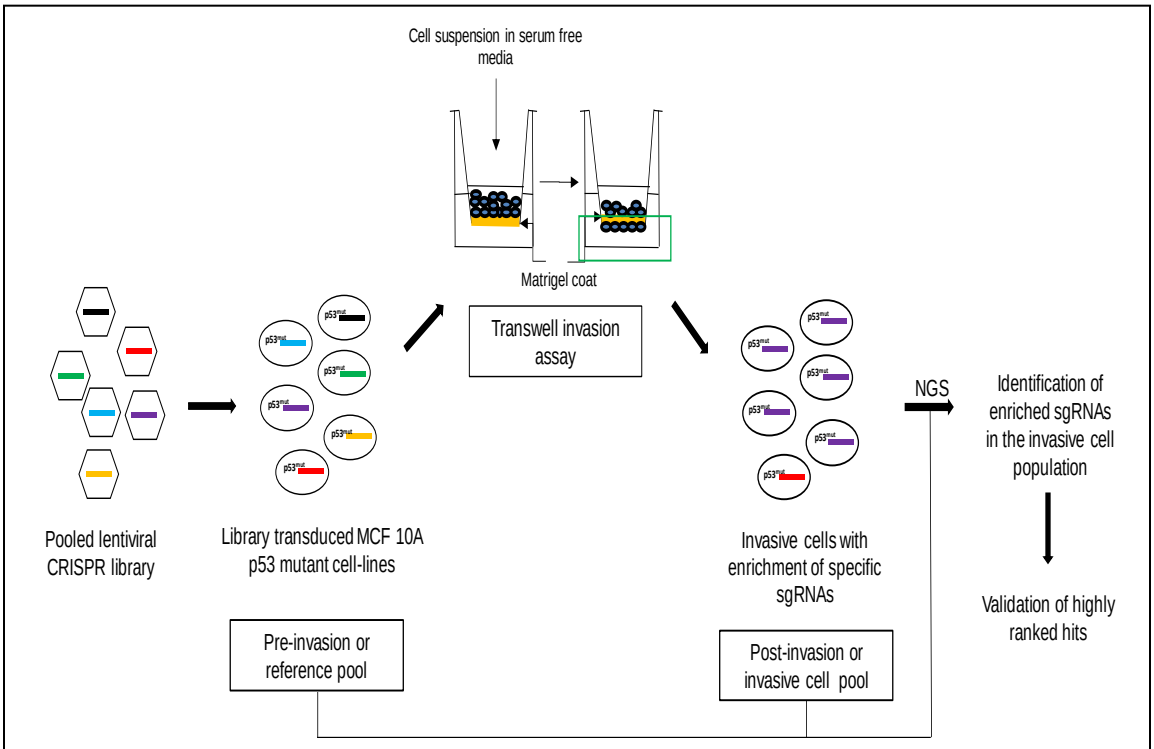


Figure 4.2. **Method for screening the co-drivers of missense mutant p53 that promote cell invasion.** The library transduced cells (p53 mut+ 1 sgRNA) were used for the invasion assay and only the invasive cells were extracted. Genomic DNA was obtained from the pre and post-invasive cells and the region containing the sgRNA was PCR amplified. The amplicon was then sequenced from the pre and post-invasive cells and enrichment of sgRNAs in the invasive pool determined.

sgRNAs in the invasive cells represented the possible co-drivers of different missense mutant p53. The top enriched sgRNAs were individually cloned into the lentiCRISPR v2 plasmid (the backbone vector of GeCKO v2.0 library) and transduced into the cells, and the increased cell invasion capacity in cooperation with mutant p53 was validated.

For the screening of unique co-drivers of missense mutant p53, we chose two cell lines with contrasting phenotypes, the least invasive Y234C and the most invasive R273C cells. We also included the MCF 10A p53 WT-OE (wildtype p53 overexpressing) cells as the control.

Method

Amplification and sequencing of the GeCKO v2.0 plasmid library

GeCKO v2.0 library obtained from Addgene was diluted to 50 ng/ul (water or TE buffer). 2 ul of the library was added to 25 ul of electrocompetent Lucigen Endura cells and electroporated into the cells. Cells were recovered with 975 ul of recovery media and then transferred to 1 ml more of recovery media. A total of four electroporations were done and the tubes rotated at 250 rpm for 1 hr at 37^{o}C. All electroporated cells were pooled (total 8 ml) and mixed well. 10 ul was added to 1 ml of recovery media and 20 ul of cells were plated. This 40,000-fold dilution of full transformation gave an estimate of the transformation efficiency and ensured full library representation. The 8 ml of transformation was distributed to 4 bioassay plates using a cell spreader and incubated for 14 hrs at 32^{o}C. The number of colonies on the diluted plate were counted and multiplied by 40,000 to get the total number of colonies. The total number of colonies were more than 3 million and thus, we proceeded with the next steps. 10 ml of LB was put onto each bioassay plate and the colonies were scraped using the cell spreader. The liquid with the scraped colonies were collected in a tube and this step repeated with another 5- or 10-ml LB. The bacteria were

centrifuged, and the bacterial pellet was weighed. Several maxi-preps were performed to extract the amplified plasmid library (each maxi-prep column can hold 0.45 gm of the bacterial pellet). This plasmid library was sequenced to verify the representation using the primers in Table 4.2.

Table 4.2. **Primer pair used for checking GeCKO v2 library representation.**

Set 3	Forward	GGCTTTATATATCTTGTGGAAAGG	U6 promoter close to the cut site
Set 3	Reverse	GACATCACTTTCCCAGTTTACC	EF-1 alpha core promoter

Preparation and sequencing of the lentiviral library

Lentivirus from the plasmid library was prepared in 6-well plates. On Day 1, 1.5 million 293Lx cells were plated in each well in 2 ml media. On Day 2, a transfection master mix was prepared containing mix 1 and mix 2. Mix 1 of total 150 ul had diluted Fugene HD in serum free DMEM media (Fugene HD ul: total ug of DNA = 6:1). Mix 2 had packaging plasmids 500 ng VSVG, 2500 ng Gag-pol and 2500 ng of lentivector in serum free DMEM media making a total of 150 ul. Mix 1 and mix 2 were combined and incubated at room temperature (RT) for 20 to 30 min. After the incubation, 150 ul of the master mix was dispensed into 2 wells each. The volume of the master mix and hence the concentration of the plasmids were increased according to the number of wells transfected. The 6-well plate was spun at 2250 rpm for 30 min and then placed in the incubator at 37^{o}C. 1% Westcodyne was used for disposing biohazard waste. On Day 3, 1 ml media was removed and replaced with 2 ml fresh media in the morning and in the evening 2.5 ml media was removed and replaced with 5.5 ml fresh media in each well. On Day 4, 5.5 ml media was collected in a 50 ml tube and stored at 4^{o}C and 3.5 ml fresh media was put into each well. Finally, on Day 5, 4 ml media was collected from each well in the same tube. The lentivirus was

concentrated using the Lenti-X concentrator (Clontech). 1 volume of concentrator was mixed with 3 volumes of virus containing media and mixed by gentle inversion. The mixture was incubated overnight at 4^{o}C and centrifuged at 1500xg for 45 min at 4^{o}C. The off-white virus pellet was resuspended with a low volume of media, immediately aliquoted and stored at -80^{o}C. The packaged plasmid from the virus was extracted using the QIAamp MinElute virus spin kit (Cat. No. 57704) as per the manufacturer's protocol. The PCR amplicons were sequenced to check the library representation in the lentivirus.

MOI determination

For determining the conditions for maintaining low MOI (multiplicity of infection) during transduction of MCF 10A p53 mutants Y234C and R273C cells with the lentiviral libraries A and B, different cell concentrations, virus volumes, selection reagent concentration and selection duration were established. 35,000 cells/well were plated in each well of 6-well plate in 2 ml on Day 1. On Day 2, cells were transduced with different volumes of lentivirus libraries A and B. 1 ml media was removed and polybrene at final concentration of 8ug/ml per well was added before addition of the virus. The virus was removed, and fresh media added on Day 3. Selection reagent (puromycin 0.7 ug/ml) was added to the cells on Day 4 and selected for 72 hrs. The un-transduced cells with or without puromycin were included as controls. The un-transduced cells with puromycin were dead (floating cells) by 72 hrs. At the end of 72 hrs, %transduction was calculated as: (total no. of transduced cells/ total no. of untransduced cells) *100. %transduced or %survival corresponding to 20%-40% ensured low MOI of 0.3 to 0.5 for transduction.

Transduction of MCF 10A p53 mutant cells with the lentivirus library

For the transduction of MCF 10A p53 mutant Y234C and R273C cell-lines, a total of 66 million and 58 million cells were transduced with the CRISPR half-library A and B respectively per cell-line at MOI=0.3. Such cell numbers were transduced in order to maintain a 250-fold

representation of each of the half-libraries. The 66 and 58 million cells were split into two batches for the ease of handling. On Day 1, cells were plated in the large cell culture T225 flasks. On Day 2, lentivirus libraries A and B were added after the addition of polybrene and on Day 3, the media in the flasks was replaced by fresh media. Puromycin (0.7 ug/ml) was added to the media on Day 4 and cells selected for 72 hrs. The cells from the T225 flasks were combined and transferred to petri-dishes and selected for at least 14 days in puromycin before using them for the invasion assay. Cells were passaged at regular intervals where half of the cells were propagated and the rest frozen as stocks.

Isolation and expansion of the invasive cells

For the extraction of invasive cells, the 6-well BioCoat Matrigel invasion chamber plates (cat # 354481) were used. 1 million cells were seeded in each Matrigel coated insert in serum free media (2ml), making a total of 12 million cells (2 plates). In the bottom well 2.5 ml media with serum was placed. After seeding, the plates were incubated at 37^{o}C and 5% CO_2 for 30 hrs for Y234C cells and 18 hrs for R273C cells. Following incubation, media from both the insert and the well were removed. The inserts and the wells were washed with 1X PBS. Using a cotton swab the cells on top of the insert, the non-invasive cells, were scraped off. 500 ul of trypsin was added to the bottom well and the insert now with only the invasive cells was placed on it. After 15 to 20 min (if the typsinization was complete, cells were completely round and floating) media containing serum was added to the bottom well to neutralize trypsin. The post-invasion or invasive cells were collected from each well in a tube and then re-plated in a new sterile 6-well plate. Once the invasive cells started to grow, they were transferred to petri-dishes. Pellets were generated for genomic DNA (gDNA) extraction and frozen vials of cells were also prepared. Genomic DNA from 5 or 12 million pre-invasion and post-invasion cells were extracted using the Blood and Cell culture DNA midi kit from Qiagen (cat # 13343) as per the manufacturer's protocol.

Sequencing of pre and post-invasion cells

The genomic region containing the sgRNA was amplified from genomic DNA of the cells using a two-step PCR method. A 700 bp region was first amplified using the following primer set 2 and PCR (PCR 1) program. PCR 1 product was purified. Using the purified PCR 1 product as the template, PCR 2 was performed to get a 300 bp amplicon. The amplicon size was verified on 1% agarose gel. This 300 bp amplicon was then sequenced (2X150 bp) to determine the enriched sgRNAs in the post-invasion pool of cells (Table 4.3).

We performed the genome wide screen for targeted loss of function of genes in MCF 10A p53 mutant Y234C and R273C expressing cells including the p53 WT-OE cell lines as the control. Sample was sequenced using the Illumina MiSeq platform. sgRNA sequences were extracted by trimming forward and reverse primer using Cutadapt, then matched to GeCKO v2.0 library for direct hit (Martin 2011). For each gene, reads from pre and post invasion samples were normalized to count per million (CPM). Post-invasion CPM was divided by pre-invasion CPM for calculation of enrichment scores. Two replicate experiments were performed with WT-OE; three with Y234C and four with R273C. We ranked the maximum enrichment score for the sgRNAs in the post-invasion cell pool to select the top hit genes.

Cloning of individual sgRNA into vector

To clone the target sgRNA sequences into lentiCRISPRv2 plasmid, two oligos were synthesized for each sgRNA (Appendix C). After the vector digestion with BsmBI, the plasmids bear an overhang that matches with that on the synthesized oligos. For the cloning of the individual sgRNAs into the lentiCRISPRv2 plasmid, the following steps were performed (Table 4.4).

Table 4.3. **Specification of 2-step PCR.** PCR 1 with primer set 2 was followed by PCR 2 with primer set 4 to obtain 300bp amplicon for NGS.

Forward primer set 2	GAGGGCCTATTTCCCATGATTCC
Reverse primer set 2	GGCCGATGCTGTACTTCTTGTCC
Template	2 ul
10 uM Forward primer	2.5 ul
10 uM Reverse primer	2.5 ul
2X Sapphire mix	25 ul
Nuclease free water	18 ul
Total	50 ul
98°C	1 min
98°C 55°C 72°C	30 cycles 15 sec 15 sec 15 sec
72°C	5 min

Forward primer set 4	CGATTTCTTGGCTTTATATATCTTGTGGAAAGGACG
Reverse primer set 4	CACGACATCACTTTCCCAGTTTACCCCGC
Template from PCR1	5 ul
10 uM Forward primer	2.5 ul
10 uM Reverse primer	2.5 ul
2X Sapphire mix	25 ul
Nuclease free water	18 ul
Total	53 ul
98°C	1 min
98°C 58°C 72°C	10 cycles 15 sec 15 sec 15 sec
72°C	5 min

Table 4.4. **Detail for cloning of individual sgRNA.**

Vector digestion and dephosphorylation 30 min at 37°C	
lentiCRISPRv2	5 ug
FastDigest BsmBI	3 ul
FastAP	3 ul
10X FastDigest buffer	6 ul
100 mM DTT freshly made	0.6 ul
Nuclease free water	Made upto 60 ul total reaction

Vector digestion for 1 hr at 55°C	
lentiCRISPRv2	5 ug
NEB BsmBI	5 ul
10X NEB buffer	6 ul
Nuclease free water	Made upto 60 ul total reaction

Oligo pair annealing and phosphorylation	
Oligo 1 100 uM	1 ul
Oligo 2 100 uM	1 ul
10X T4 ligation buffer NEB	1 ul
T4 PNK NEB	0.5 ul
Nuclease free water	Made upto 10 ul total reaction

Oligo pair annealing	
Oligo 1 100 uM	1 ul
Oligo 2 100 uM	1 ul
Nuclease free water	Made upto 10 ul total reaction

Ligation reaction at RT for 10 min	
BsmBI digested plasmid dephos/ no dephos	50 ng
Diluted oligo duplex phos/ no phos	1 ul
2X Quick ligase buffer NEB	5 ul
Nuclease free water	Made upto 10 ul total reaction
Quick Ligase NEB	1 ul

1) Digestion of lentiCRISPRv2 plasmid. The vector was either digested and dephosphorylated or simply digested. 2) Gel extraction of cut plasmid. BsmBI digestion resulted in a 2 kb filler fragment and a larger band. The digested products were checked on 1% agarose gel run for 27 min at 120 V. For cloning, only the larger band was extracted from the 1% GTG agarose gel run for 27 min at 120 V. The extracted band was purified using the QIAquick gel extraction kit as per manufacturer's protocol. 3) Annealing each pair of oligos. The oligo pairs were either annealed and phosphorylated or simply annealed. For phosphorylation and annealing of the oligos, the thermocycler was set as 37^{o}C for 30 min, 95^{o}C for 5 min and then ramp down to 25^{o}C at 5^{o}C/min. For simply annealing the oligos, the setting was 95^{o}C for 5 min and then ramp down to 25^{o}C at 5^{o}C/min. The annealed oligos were diluted (1:200) in sterile water. 4) Ligation reaction. The dephosphorylated cut plasmid was ligated with the phosphorylated oligos and the simply cut plasmid was ligated with the annealed oligos without phosphorylation. As a negative control we used water instead of the oligos with the digested vector. 5) Transformation into Stbl3 bacteria. 25 ul Stbl3 cells were mixed with 5 ul ligation reaction by gentle flicking. Cells were incubated on ice for 15-20 min. For transformation, cells were given heat shock at 42^{o}C for 30-40 s and then immediately put on ice for 2 min. 200 ul of SOC media was added to the cells and shaken at 250 rpm for 1 hour. All the cells were plated on agar plates with appropriate selection reagent and incubated at 37^{o}C overnight. After 16 to 18 hrs if colonies appeared, 3 to 4 random colonies were picked for 5 ml overnight cultures. The agar plates were then stored at 4^{o}C. Plasmids were extracted from the overnight cultures and colony PCR was performed. The region of the plasmid containing the cloned sgRNAs was PCR amplified and the sequence checked with Sanger sequencing.

Transduction of MCF 10A p53 mutant cells

Lentivirus packaged with cloned sgRNA for *GCNT1, FGD6, ADAM15, PACS1* and *SDF4* were prepared in 6-well plates. 1.5 million 293Lx cells were plated in each well in 2 ml media. A transfection master mix was prepared the next day containing mix 1 and mix 2. Mix 1 of total 150

ul had diluted Fugene HD in serum free DMEM media (Fugene HD ul: total ug of DNA = 6:1). Mix 2 had packaging plasmids 500 ng VSVG, 2500 ng Gag-pol and 2500 ng of lentivector with the appropriate cloned sgRNA in serum free DMEM media making a total of 150 ul. Mix 1 and mix 2 were combined and incubated at room temperature (RT) for 20 to 30 min. After the incubation, 150 ul of the master mix was dispensed into 2 wells each. The 6-well plate was spun at 2250 rpm for 30 min and then placed in the incubator at 37^{o}C. On the following day, 1 ml media was removed and replaced with 2 ml fresh media in the morning and in the evening 2.5 ml media was removed and replaced with 5.5 ml fresh media in each well. The next day 5.5 ml media was collected in a 50 ml tube and stored at 4^{o}C and 3.5 ml fresh media was put. Finally, on the last day 4 ml media was collected from each well in the same tube. The lentivirus was concentrated using the Lenti-X concentrator (Clontech).

MCF 10A p53 mutant Y234C cells were transduced with lentiviruses with sgRNA for GCNT1, FGD6, ADAM15 and R273C cells with lentiviruses containing sgRNA for PACS1 and SDF4. For transduction 75,000 cells/well in 2 ml media were plated in 6-well sterile plates. On the following day, 1 ml media was removed from each well and desired amount of lentivirus was added after the addition of polybrene. Next day media was replaced by fresh media. Puromycin (0.7 ug/ml) was added to the media on the following day and cells selected for 72 hrs. The cells were transferred to petri-dishes and selected for at least 14 days in puromycin before further use. Cells were passaged at regular intervals and frozen stocks were also made.

T7 endonuclease assay and MiSeq of the gene locus

Genomic DNA (gDNA) was extracted from the cells and the sgRNA target locus was PCR amplified. gDNA from MCF 10A Y234C and R273C cells (not transduced with the library) were used as negative controls and water as no template control (NTC). Detail of the PCR (Table 4.5): The PCR products were run on 1% agarose gel to verify the size and amplification. The PCR products were then purified, and their concentrations determined. The PCR products were then

set for the annealing reaction. 1 ul of T7 endonuclease I enzyme (NEB cat # M0302) was added to 19 ul annealed PCR products and incubated for 15 min at 37°C. The reaction was stopped with 1.5 ul of 0.25 M EDTA. One may also consult the protocol provided in the NEB website. The fragmented PCR products were analyzed on 10% TBE gel run in 0.5X TBE buffer for 70 min at 120V. The gel was then stained in 75 ml 0.5X TBE buffer with 0.7 ul EtBr for 20 min.

For sequencing of the targeted locus, the locus was PCR amplified as per the Table 4.6, purified and submitted for MiSeq 2x150 bp run. Paired-ended sequences were aligned to the reference human genome (GRCh38.p92/hg38) using Bowtie2 (Bowtie2 2.1.0) (Langmead B, 2012). A deletion was called if a deletion site was within ± 5bp apart from the sgRNA targeting site. Igvtools (igv_2.4.13) was used for data visualization (Helga 2013).

Table 4.5. **PCR and annealing step for T7 endonuclease assay.**

PCR component	50 ul reaction	Step	Temperature	Time
Sapphire 2X master mix	25 ul	Initial denaturation	98°C	30 s
10 uM forward primer	2.5 ul	35 cycles	98°C	5 s
10 uM reverse primer	2.5 ul		50-72°C depending on the Tm of primers	10 s
			72°C	20 s
Template DNA	100 ng in total	Final extension	72°C	2 min
Nuclease free water	Upto 50 ul	Hold	4°C	

Component	19 ul annealing reaction	Step	Temperature	Ramp rate	Time
DNA	200 ng	Initial denaturation	95°C		5 min
10X NEBuffer 2	2 ul	annealing	95-85°C	-2°C/sec	
			85-25°C	-0.1°C/sec	10 s
Nuclease free water	Upto 19 ul	Hold	4°C		

Table 4.6. **PCR amplification of target locus for MiSeq.**

PCR component	50 ul reaction	Step	Temperature	Time
Sapphire 2X master mix	25 ul	Initial denaturation	98°C	30 s
10 uM forward primer	2.5 ul	35 cycles	98°C	15 s
			50-72°C depending on the Tm of primers	15 s
10 uM reverse primer	2.5 ul		72°C	15 s
Template DNA	2 ng/ul in total	Final extension	72°C	5 min
Nuclease free water	Upto 50 ul	Hold	4°C	

Western blot

Cells were grown in 6-well plates for the cell lysate preparation. Cell lysates were made using the RIPA lysis buffer containing 50mM Tris, 150mM NaCl, 1% IGEPAL CA-630, 0.1% sodium azide (Sigma #S8032), cOmplete mini protease inhibitor (Sigma #04693124001), 200 µM sodium fluoride and 200 µM of sodium orthovanadate. Cells when confluent were scraped, incubated on ice for 10 minutes and centrifuged at 10,000 rpm for 10 min to collect the clear lysate. Lysates were quantified using the Pierce BCA kit (cat # 23225). Samples were run on 4-20% TGX gels from BioRad (#567-1093) and blotted onto 0.45µm PVDF membrane (GE Healthcare #10600023) using the BioRad semi-dry transfer system. Primary antibody for GCNT1 (Thermo PA5-12146), FGD6 (Thermo PA5-48123), ADAM15 (Sigma Aldrich HPA011633), PACS1 (Invitrogen PA5-72865) and SDF4 were diluted to 1:500 and secondary HRP-linked anti-mouse (1:5000) or anti-rabbit (1:3000) antibody used to visualize the protein bands. The blots were imaged on the Fluorchem FC3 from AlphaInnotech.

Cell invasion assay in 24 well plates

Cells (70-80% confluent) were trypsinized and used for the transwell invasion assay (24 well). Before the experiment, the Matrigel coated invasion plate (Corning cat # 354480) was thawed with 200 ul serum free media in each well for rehydrating the matrigel layer at 37^{o}C for 1 hr. 2.5x10^{5} cells in 200ul serum free media were seeded in the Matrigel coated insert. 750 ul of serum containing media was added to the bottom wells. For the Y234C cells with the co-drivers, the invasion plates were incubated for 24 hrs and for R273C cells with co-drivers incubation time was 18 hrs. After incubation, cells were fixed with 4% paraformaldehyde for 2 min at RT, permeabilized by adding 500ul of 100% cold methanol (20 min at RT) and stained with 0.5% crystal violet dye (15 min at RT). A cotton swab was used to remove any non-invasive cells. The invasive cells were imaged under a microscope at 4X and 10X magnification.

Results

Representation of GeCKO v2.0 plasmid and lentiviral library

The chapter 3 showed the proof-of-principle for ‘co-driver’ of missense mutant p53 where known deletion of PTEN by CRISPR-Cas9 increased invasive potential of the weakly invasive MCF 10A p53 mutant G245S, R273H and Y234C cells. To scale up the pipeline for a genome wide screen for the co-drivers of missense mutant p53, a CRISPR based library, GeCKO v2.0, was purchased from Addgene. Functional genomic screening for the co-drivers enabled identification of the perturbagen (sgRNA targeting a gene in this case) that generated the invasive phenotype.

The GeCKO v2.0 library was transformed, expanded, and purified to have sufficient concentration and amount of the plasmid DNA to prepare the lentivirus library with complete coverage of all

sgRNAs. The library representation was verified using NGS at each stage, after amplification of the plasmid library and after preparation of the lentivirus library. The 300 bp amplicon obtained from PCR amplification of plasmid half-library A and B was purified and sequenced (Table 4.7). For both the half-libraries A and B the representation in terms of the total number of sgRNAs was 99.8% and 99.7%, respectively. This ensured the presence of sgRNAs in half-libraries A and B targeting 99% of all human protein-coding genes. Using this plasmid library, the lentiviral library was generated. The two half-libraries A and B were always kept separate and never pooled. From the sequencing of the lentivirus half-libraries A and B, 83% and 91% of sgRNAs, respectively, were packed in the lentivirus library. But, collectively, the two lentiviral libraries contained sgRNAs targeting more than 99% of the human genes (Fig 4.3). These showed that the library representation and the sgRNA diversity were maintained during both GeCKO v2.0 library amplification and lentivirus production.

Table 4.7. **Coverage of plasmid half libraries A and B from NGS.** Below is the primer pair used to obtain the amplicon for NGS.

	Total genes	Found in NGS	% representation	Total gRNA	Found in NGS	% representation
Half-library A	20915	20669	**98.8**	63950	63807	**99.8**
Half-library B	19050	18834	**98.9**	56869	56694	**99.7**

Set 3	Forward	GGCTTTATATATCTTGTGGAAAGG	U6 promoter close to the cut site
Set 3	Reverse	GACATCACTTTCCCAGTTTACC	EF-1 alpha core promoter

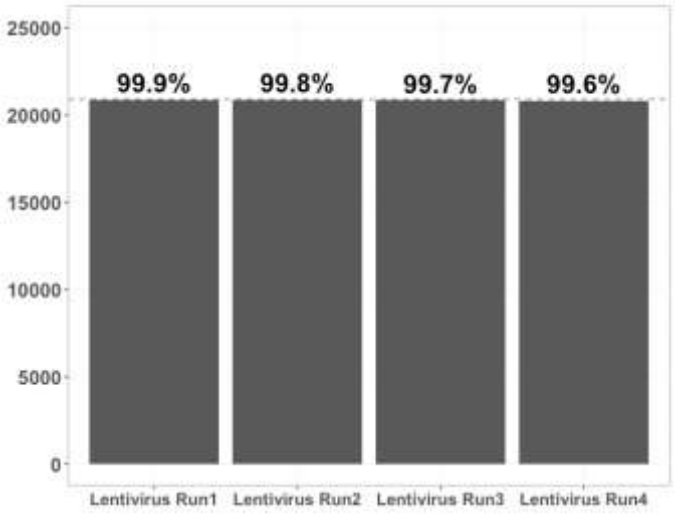


Figure 4.3. **Coverage of the genes by the lentivirus library determined by NGS.**

Transduction of cells with lentiviral GeCKO v2.0 library at low MOI

After the lentivirus library preparation, the optimal conditions for transduction of the MCF 10A p53 mutant cells were established. The conditions were carefully worked out to ensure that a given cell was infected by a single virus particle and thus, only one sgRNA integrated into the genome. This ensured that the p53 mutant cells had only one additional mutation (1 sgRNA integration per cell).

Multiplicity of infection (MOI) is the ratio of number of virus particles and number of host cells in an infection. MOI was determined and adjusted by changing the concentration of the virus and the cells. To allow transduction of only one sgRNA per cell, MOIs between 0.3 and 0.5 were considered, which corresponded to the percent survival or the percent transduction between 20% and 40%. The cell concentration and the lentivirus volumes were thus adjusted to yield 20% to 40% survival of transduced cells after transduction with the GeCKO v2.0 library and puromycin selection (Fig 4.4). In order to maintain a 250-fold coverage for each sgRNA, 66 million and 58 million cells were separately transduced with the lentiviral half-library A and B, respectively. The

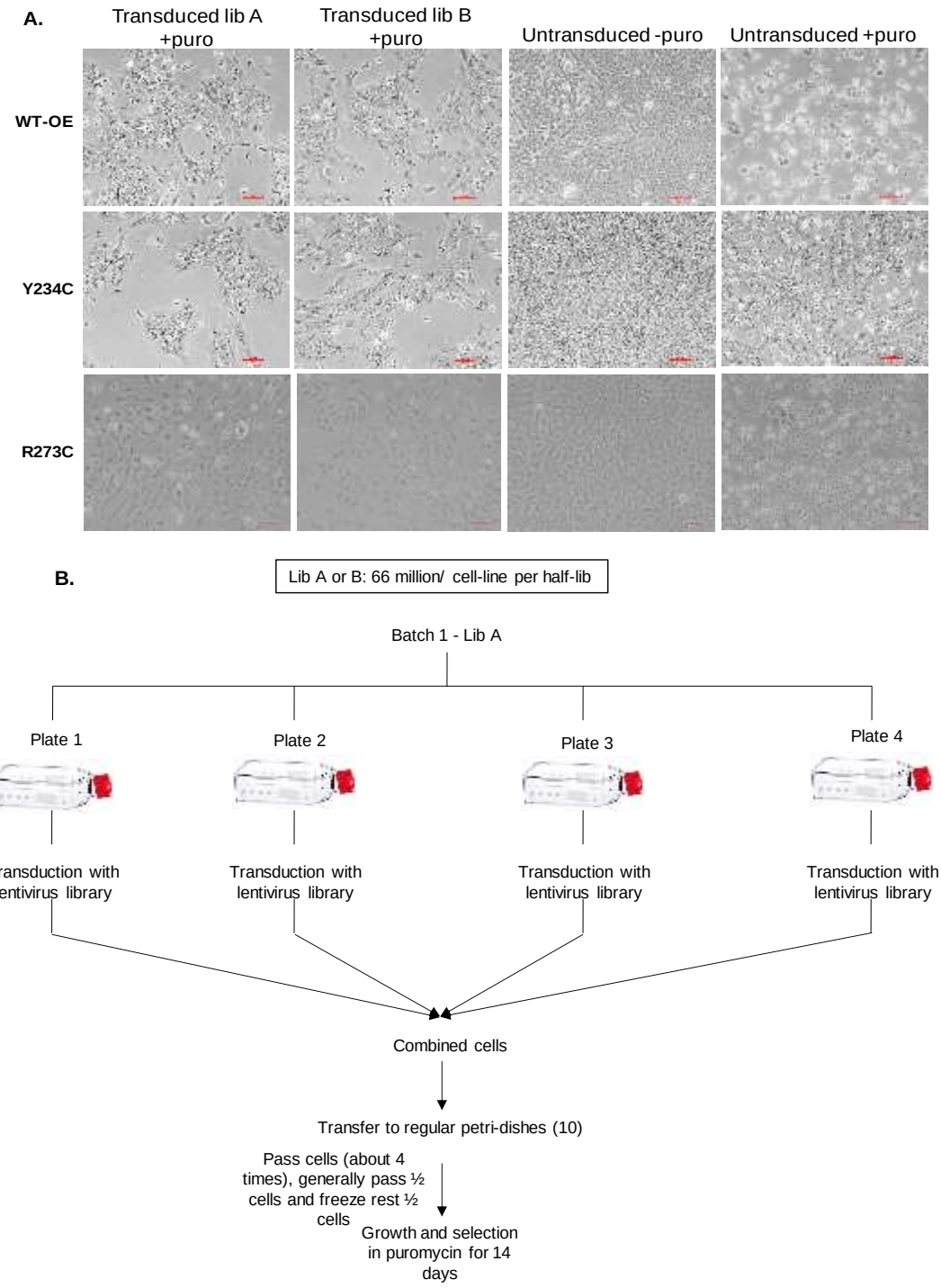


Figure 4.4. **Cells transduced at low MOI.** A. Low MOI of 0.3 for transduction corresponded to 20-40% cell survival after selection. Untransduced cells without or with the selection reagent were included as the controls. B. Independent transduction to ensure complete coverage by the library for co-driver screening.

cells were selected with puromycin for 2 weeks before using them for phenotypic selection by the invasion assay.

A pilot study revealed at least 3 independent screens were needed to test all genes as possible co-drivers of missense mutant p53

Although, we calculated the mathematical cell number required to achieve 250x coverage of every sgRNA, the actual coverage might be influenced by various biological and technical factors. Thus, to further ensure a complete coverage of genes during the screen, we then examined how many independent batches of transduced cells were needed to test and screen all human genes. One of the major biological factors that can affect the coverage is the fraction of genes critical for cell survival, such as housekeeping genes, or cell proliferation. Around 30% of the human protein-coding genes are considered essential, and sgRNAs that ablate the function of the essential genes are thus, unlikely to be detected in the transduced cells after puromycin selection. In addition, due to the technical factors, a small fraction of sgRNAs might be lost during the packaging into the lentivirus or during transduction of cells. Therefore, aiming to find the optimal experimental setup that can maximize the number of protein-coding genes to be screened, multiple rounds of independent transduction, puromycin selection, and sequencing were performed on 3 cell lines (WT-OE, Y234C, and R273C), and individual and cumulative empirical coverages of the replicates were calculated.

As shown in Fig 4.5 the first batch of transduced WT-OE cells (Fig 4.5A, “run 1”) had only a small fraction of sgRNAs (~23,300 or 21% of all 110,819 sgRNAs for protein-coding genes in Lib A and B combined), but, when mapped to the genes, the sgRNAs covered 87% of all protein-coding genes (18,262 out of 20,700) with 1.27 sgRNAs/gene ratio. In Y234C and R273C cells, we observed the gene coverage of 83% and 57% with the sgRNA/gene ratio of 2.12 and 1.44, respectively. When we calculated the cumulative coverage over the multiple batches of transduced cells, combinations of 2 or 3 batches could cover 90% of all protein-coding genes with

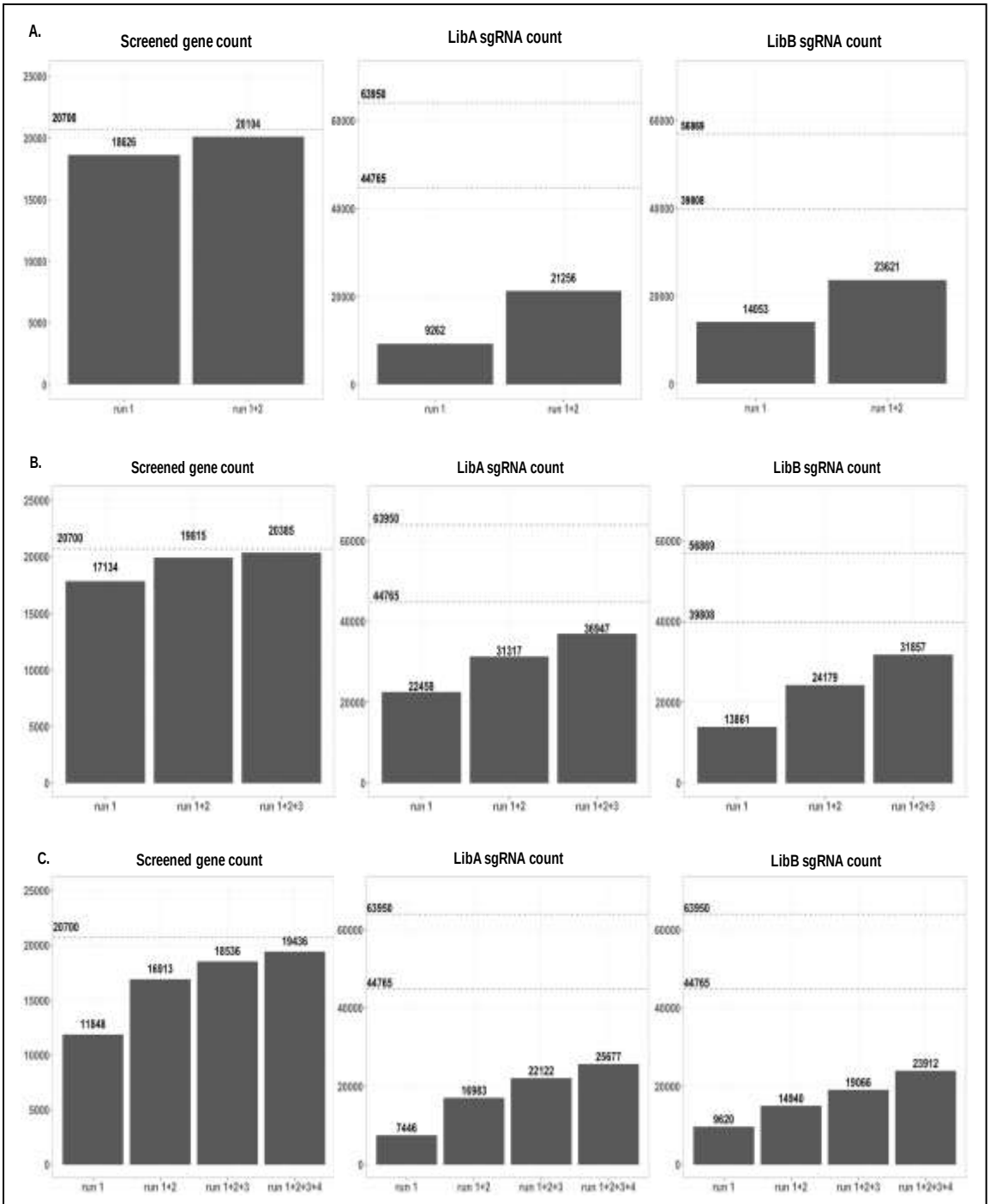


Figure 4.5. **sgRNA counts and gene coverage from separate NGS runs of library transduced pre-invasion cell pool (reference pool).** A. 2 sequencing runs of WT-OE-libA and libB cells. B. 3 sequencing runs of Y234C-libA and libB cells. C. 4 sequencing runs of R273C-libA and libB cells. Although all the sgRNAs were not fully covered in the reference pool, there was full coverage of all the 20,000 genes targeted by the sgRNAs. At least 3 independent screens were needed for screening all genes for the co-drivers.

the sgRNA/gene ratio greater than 2.0. Although different p53 mutant-expressing cells showed different library transduction efficiency likely due to distinct set of essential genes and different cellular states, these results demonstrated that three independent screens were sufficient to perturb the majority of protein-coding genes with 2 or more sgRNAs and test the phenotypic influence for identification of invasion promoting co-drivers of missense mutant p53.

Screening for co-drivers of missense mutant p53 that drive cell invasion

We screened a total of 48 million (combining from all 3 batches) library transduced Y234C cells (24 million from Y234C-libA and 24 million of Y234C-libB) for co-drivers promoting cell invasion. For library transduced R273C cells, a total of 96 million cells (48 million cells from R273C-libA and R273C-libB each) were screened for co-drivers. WT-OE-libA and libB cells were used as the control.

The library transduced cells of weakly invasive p53 mutant Y234C (Y234C-libA and Y234C-libB) were incubated for 30 hrs while the library transduced cells of more invasive p53 mutant R273C (R273C-libA and R273C-libB) were incubated in the Matrigel coated transwell invasion chambers for 18 hrs. The library transduced cells demonstrated increased cell invasion compared to the non-transduced Y234C and R273C cells indicating that the gene deletion by sgRNA enhanced cell invasion (Fig 4.6).

To identify the co-drivers of missense mutant p53 that enhanced cell invasion, the invasive cells that managed to cross the matrigel barrier were extracted. As only 10 to 15% of transduced cells invaded, the number of extracted invasive cells were too low to proceed with genomic DNA extraction for NGS. Hence, the extracted invasive cells were plated into new sterile 6-well plates and grown for 5 to 7 days. Genomic DNA (gDNA) was then isolated from 5 to 12 million post-invasion cell population.

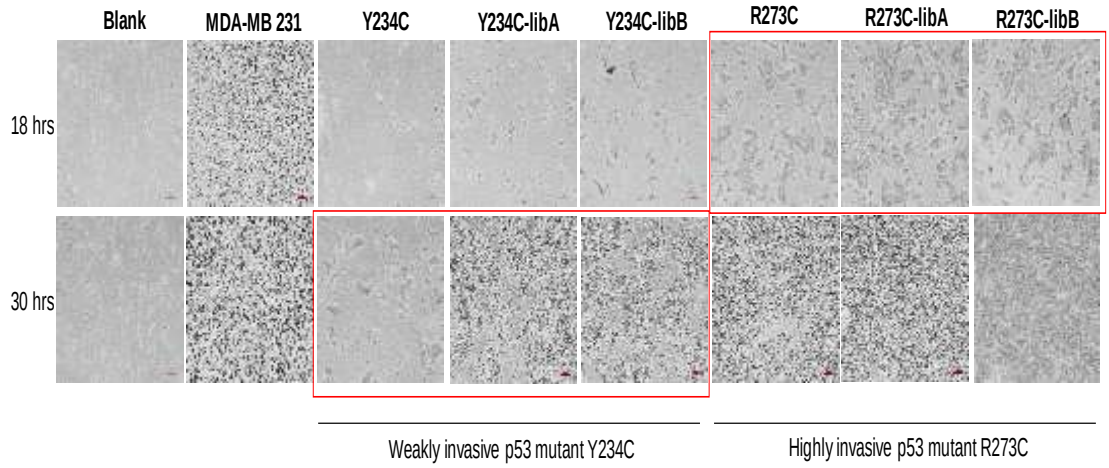


Figure 4.6. **End point determination for invasion assay for screening of co-drivers**. For the invasion assay 1 million cells were seeded in each insert per cell-line. Cells were then stained with crystal violet. LibA and libB transduced cells (for both Y234C and R273C) resulted in increased invasion of cells.

We used the MCF 10A p53 WT-OE-libA and WT-OE-libB cells as the controls to filter out the co-driver candidates that are common to the WT p53 protein. After the amplified sgRNA cassettes were sequenced, counted, and normalized to CPM (counts per million reads), to assess the relative enrichment of sgRNAs before and after invasion-based selection, post-invasion CPM was then divided by pre-invasion CPM to calculate the enrichment scores. Among the replicate screens, we selected the maximum enrichment score for each sgRNA in the post-invasion cell pool to identify the top hit genes.

The post-invasion cell population of WT-OE cells had a limited number of highly abundant (*i.e.*, CPM >1,000 or >0.1% of total reads) sgRNAs (Fig 4.7), and only a few (N=3) genes showed the fold enrichments greater than 1,000 folds over the pre-invasion population (Fig 4.8). In contrast, in the invasive pool of Y234C cells, many selected sgRNAs were abundant above 1,000 CPM up to 30,000 CPM (or >3% of total reads) (Fig 4.8), including sgRNA for *PTEN* and several known cancer related genes such *EPHB2*. Similarly, in the post-invasion cell pool of R273C-libA and R273C-libB, several highly abundant sgRNAs were detected compared to that of control cells, most notably for the *PACS1* sgRNA that occupied up to 80% of total reads (Fig 4.8). Many known

cancer genes (*FUS*, *EGLN1*) were also enriched in the R273C-libA and libB post-invasion cell pool. In addition, compared to WT-OE and Y234C post-invasion cell pool, the fold enrichments of sgRNAs in R273C post-invasion cell pool was much larger by an order of magnitude (Fig 4.8). Interestingly, the set of enriched sgRNAs were cunique to each p53 missense mutant Y234C and R273C as well as the wildtype p53.

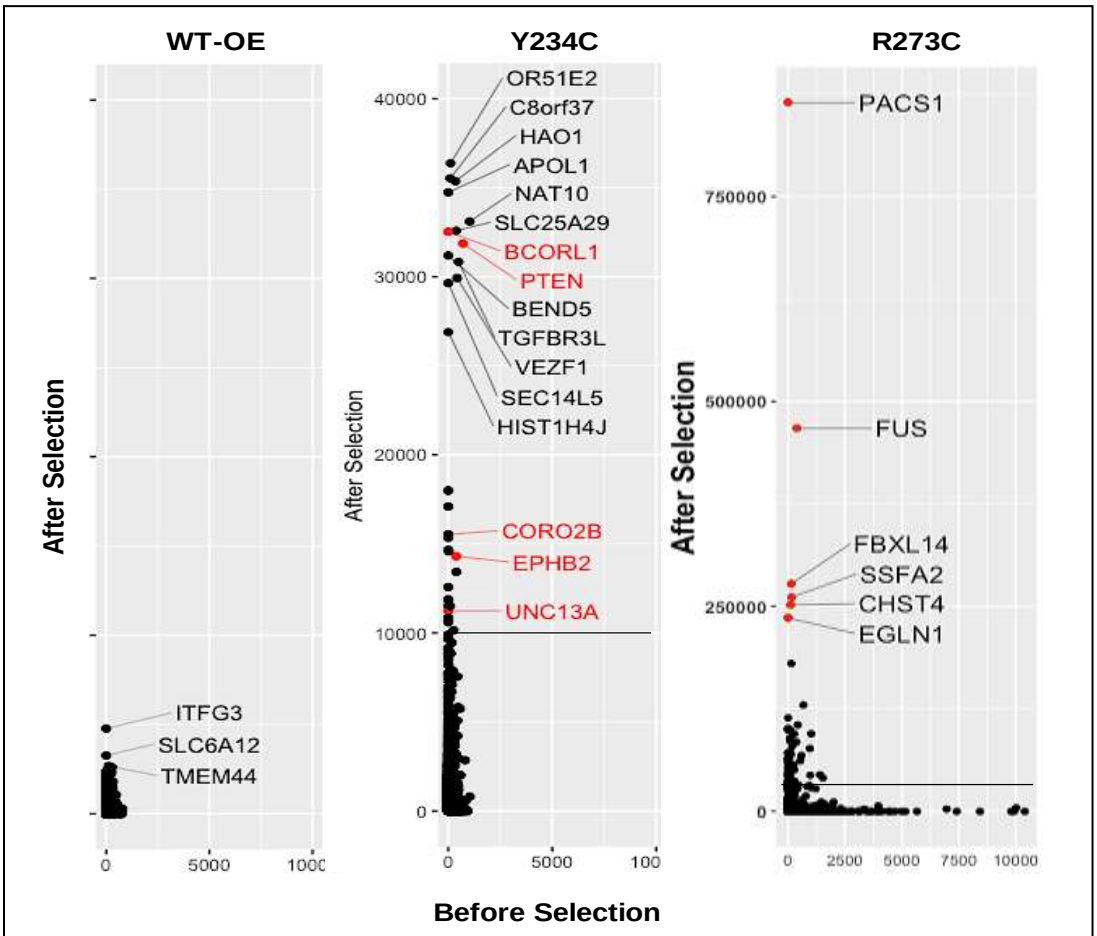


Figure 4.7. **The normalized read counts of sgRNAs in pre- and post-invasion (invasive) cell pools.** The X and Y axes depict the CPM values of sgRNAs in the pre-invasion and post-invasion cell pools, respectively.

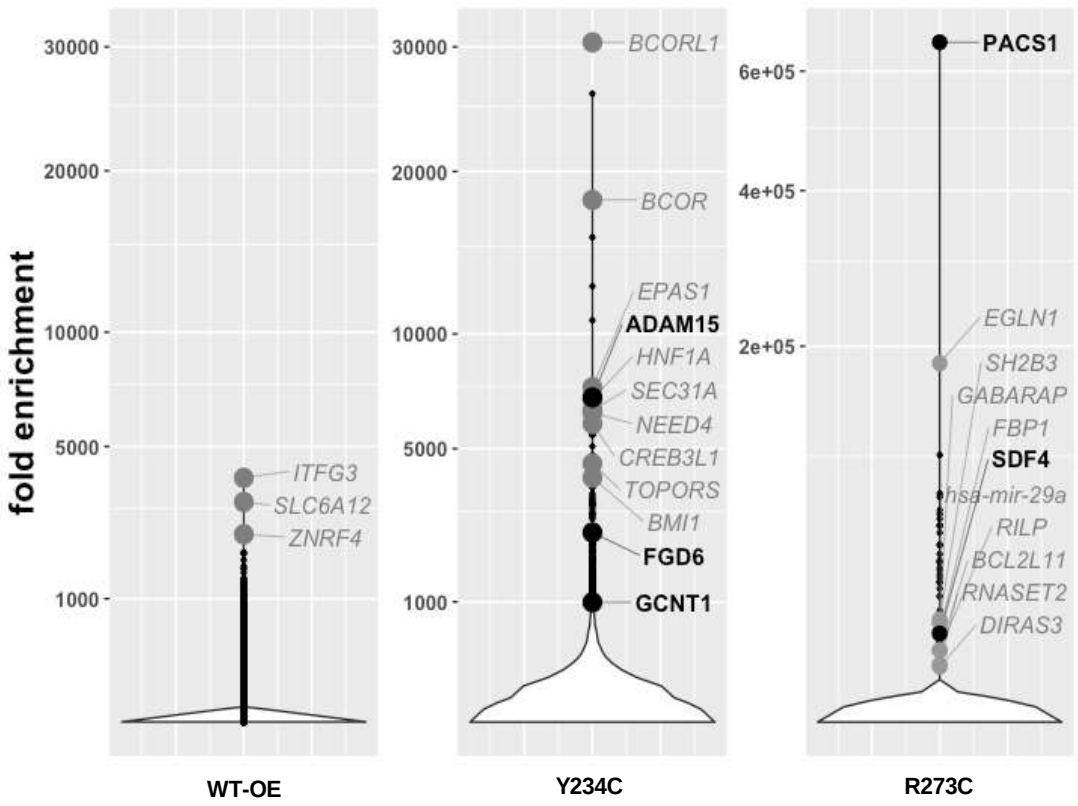


Figure 4.8. **Fold enrichments of sgRNAs in invasive populations of WT-OE, Y234C and R273C cells over the reference populations.** Genes marked with grey are known tumor suppressors among top 100 enriched genes. Black are the ones chose for validation. It is to be noted that the scales are not the same.

Analysis of the top co-driver candidates

Known p53 related genes such as *CDKN1A* and *PIK3CA* were largely enriched in Y234C (463- and 79-fold enriched, respectively), but such pattern was not observed in R273C cells. There was no correlation in enrichment scores between Y234C and R273C cells, and top enriched genes were entirely unique to each cell, indicating that a distinct set of co-drivers is needed for cell invasion. To select the top hits from the screens for functional validation, we ranked the hits based on several criteria. The top hits were first selected based on the number of times the candidate sgRNAs identified as hits (>1,000-fold enrichment) in the screenings. For example, all

the selected top hits showed up at least two or more times in the independent screens. We then examined the targeting site of the sgRNA in the gene (*i.e*., if sgRNA targets the N-term or an important functional domain of a protein), co-occurrence of somatic mutations with *TP53* mutations in patients or other experimental data sets, and its relevance to the process of cell invasion, migration and cytoskeletal organization. We also included the data from RNA seq of the p53 mutant cells for correlation between gene expression and the invasive phenotype. For example, the GO term enrichment analysis showed that apoptosis and proliferation related genes were significantly enriched among the top 100 genes from the Y234C and R273C screens. As mentioned above, the total number of extracted invasive cells were very low for any downstream processing and thus, cells were expanded for a few days, which might lead to enrichment of sgRNAs for apoptosis and cell proliferation related genes. Thus, we carefully chose the target genes likely to be associated with the process of cell invasion. Finally, though there are a total of 6 sgRNAs (3 sgRNAs in libA and 3 in libB) in the library targeting each gene, we chose to validate the only sgRNA sequence that appeared enriched in the multiple screens.

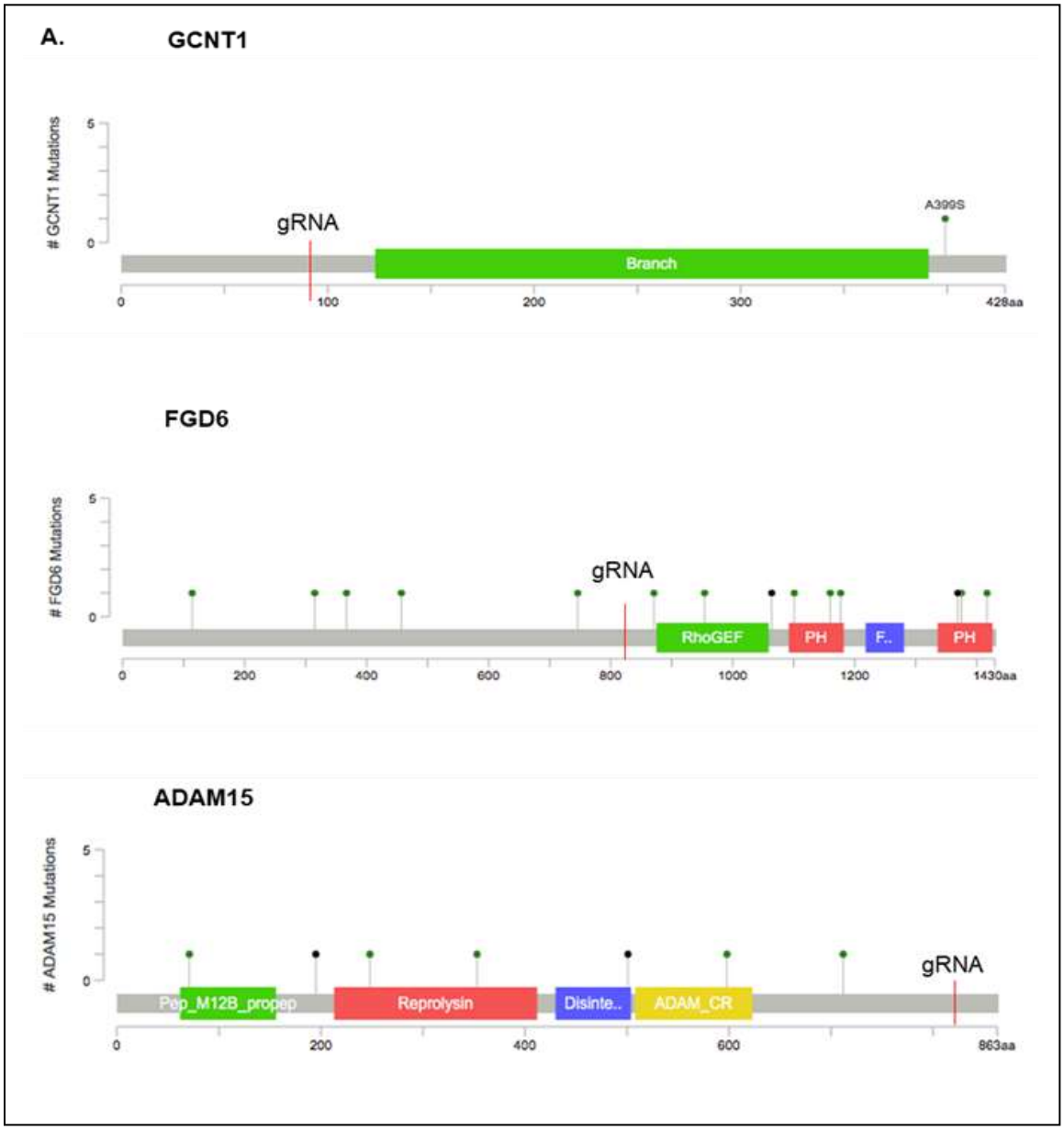
For functional validation, we chose three top hits *GCNT1*, *FGD6* and *ADAM15* for Y234C and *PACS1* and *SDF4* for R273C from the integrated analysis. These results implied that the acquisition of aggressive phenotype required different sets of co-driver mutations for different p53 missense mutations.

GCNT1 is the glucosaminyl (N-acetyl) transferase 1 and forms Gal beta 1-3 GalNAc and the core 2 O-glycan branch. It has a role in sphingolipid metabolism and cell adhesion. GCNT1 is highly expressed in prostate, endometrial cancer and colorectal cancer but is downregulated in breast cancer (Hagisawa 2005, Miyamoto 2013, St Hill 2009, Solatycka 2012). The enriched sgRNA in the screens targeted the N-terminus of the protein, which could effectively knock-out the gene. *FGD6* is the FYVE, RhoGEF and PH domain containing 6 gene. The RhoGEF domain mediates the GDP-to-GTP exchange of the Rho/Rac/Cdc42 GTPases involved in the cell migration process. The PH or the Plecstrin homology domain is generally a part of cytoskeletal and signal

transduction proteins, binds to phosphoinositide lipids, and thus functions in membrane targeting. FGD6 has a proposed role in chemotaxis of cells, actin cytoskeleton organization, filopodium assembly, regulation of GTPase activity and cell shape (NCBI database). The selected sgRNA targeted the region close to the RhoGEF domain, which thus, likely would result in a non-functional protein. The other target *ADAM15* is a metallopeptidase, a transmembrane glycoprotein involved in cell adhesion. It has a zinc binding metalloprotease domain, a disintegrin like domain and an EGF like domain. It interacts with integrin and Src family protein tyrosine kinases indicating its potential function in cell-cell adhesion and signaling. It is known to be upregulated in aggressive prostate cancer and bladder cancer (Burdelski 2017, Hiles 2016), but not much is known about its role in breast cancer. Interestingly, the enriched sgRNA targeted the C-terminal end of the ADAM15 protein instead of the catalytic zinc-binding metalloprotease domain or the N-term (Fig 4.9A).

The sgRNA targeting PACS1 was highly enriched and appeared repeatedly in the post-invasion cell pool of library transduced R273C. *PACS1*, the phosphofurin acidic cluster sorting protein 1, has a putative role in the trans-Golgi network. In humans, it is suggested to be a regulator of BAX/BAK oligomerization and downregulation of MHC-1 molecules to trans-Golgi network promoting escape from immune surveillance (Brascchio 2017). Though PACS1 does not have a suggested direct role in the process of cell invasion, we chose to validate it as it was the most highly enriched in the post-invasion cell population. Its possible role in protein sorting may indirectly help the process of cell motility by recycling of integrins and other cell adhesion proteins (Wan 1998). The sgRNA targeted the N-terminus of the PACS1 protein (exon 3 of 24 exons). The other hit, SDF4, is the stromal cell derived factor 4. It is a calcium binding protein that may regulate calcium dependent cellular activities and localizes to the Golgi lumen. It has a role in controlling cell proliferation and the secretory cargo (Crevenna 2016). SDF4 is expressed in mammary tissues and cells, and its reduced level has been associated with poor clinical outcome in breast cancer (Kang 2009). The selected gRNA targeted a region close to the calcium binding EF hand motif of the protein and may affect the calcium binding activity (Fig 4.9B).

It is striking to see a list of completely different enriched sgRNAs obtained from the screening of Y234C and R273C library transduced cells for increased cell invasion. The presence of uniquely different sgRNAs in the post-invasion pool of the Y234C and R273C library transduced cells demonstrated the context dependency of the mutations and their collaboration with separate 'co-drivers' to drive the process of cell invasion.



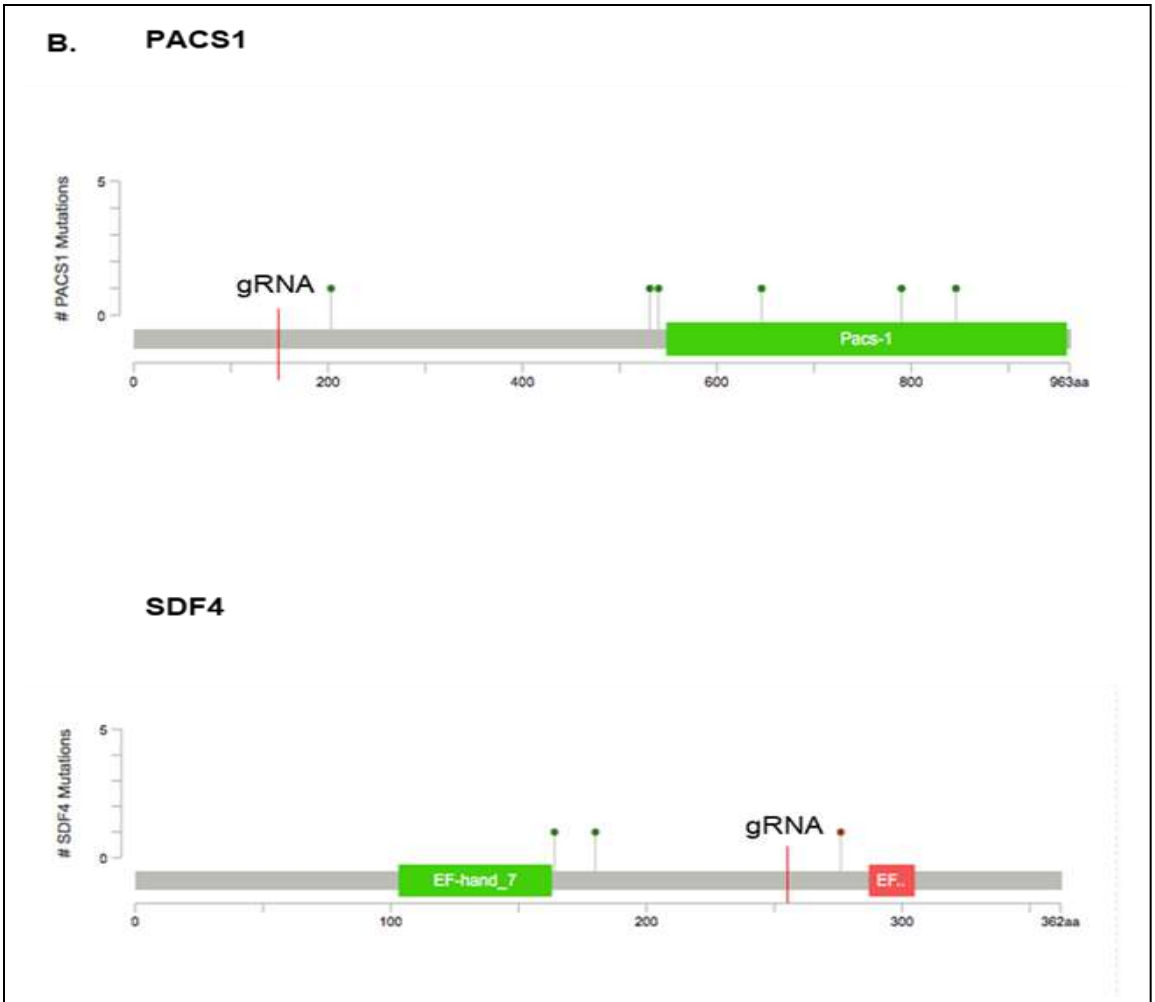


Figure 4.9. **The sgRNA target sites on top hits from the two screens.** A. The top 3 selected genes from Y234C library transduced post-invasion cell pool. The red vertical line is the position at which the sgRNA targeted in each gene. B. The top 2 selected genes from R273C library transduced post-invasion cell pool. The red vertical line is the position at which the sgRNA targeted in each gene.

Validation of the co-driver candidates

The highly enriched sgRNAs targeting *GCNT1, FGD6* and *ADAM15* as obtained from the post-invasion cell pool of Y234C cells and the ones targeting *PACS1* and *SDF4* obtained from post-invasion cell population of R273C cells were cloned individually into the lentiCRISPRv2 plasmid (backbone of the GeCKO v2.0 plasmid library). The individual sgRNAs were synthesized as

oligos, annealed and ligated with the cut lentiCRISPRv2 plasmid, which were then transformed into Stbl3 bacteria and plated. The cloning of the individual sgRNAs was verified by colony PCR and confirmed with Sanger sequencing. The sequence verified plasmids were used to prepare the lentivirus and transduce the corresponding MCF 10A p53 mutants Y234C and R273C cells. All the transduced cells were selected with puromycin for 2 weeks before further characterization.

As observed before (chapter 3), the efficiency of genome editing by CRISPR system is low, and the sgRNA transduced cells thus, represented a mixed population of cells with and without gene editing. To isolate the true CRISPR edited cells, we used the invasion assay to enrich the cells with phenotype-changing deletion. We did two consecutive rounds of invasion assay to isolate the true CRISPR edited clones, and cell lysates were prepared from the pre-invasion and post-invasion cell population of Y234C-GCNT1, Y234C-FGD6 and Y234C-ADAM15. A decrease in the protein level of GCNT1, FGD6 and ADAM15 was observed in the western blots in the post-invasion cell compared to pre-invasion cell population (Fig 4.10A).

To detect the efficiency of the gene editing by the CRISPR-Cas9 system, the T7 endonuclease I assay was performed, which recognizes and cleaves mismatched DNA. The respective sgRNA targeting loci were first PCR amplified from the genomic DNA. These PCR amplicons were rehybridized slowly and annealed products were digested with T7 endonuclease I enzyme. The presence of DNA fragments on 10% TBE gel indicated gene editing made by the CRISPR-Cas9 system (Fig 4.10B). To confirm gene editing at the genomic loci, the individual gene locus was PCR-amplified and sequenced by NGS (Fig 4.10C). The sequencing results showed that 98% of the cells carried indels at the *GCNT1* locus, 39% at FGD6 locus and just 4.8% at the *ADAM15* locus which might be due to distant deletion from the target site. CRISPR editing by small indels at the extreme C-term may result in inefficient transcription or unstable mRNAs with no protein subsequently.

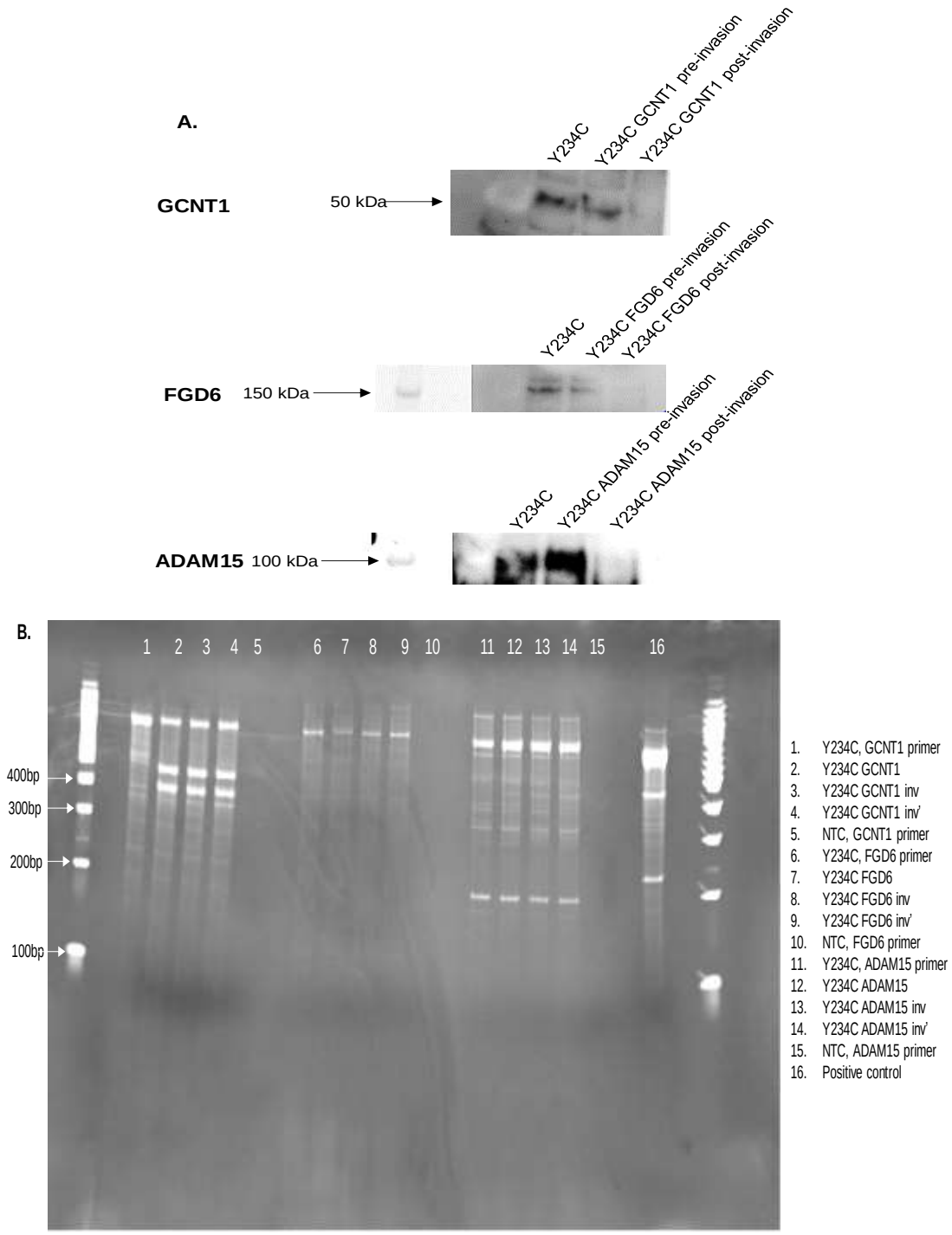


Figure 4.10. **Validation of selected co-drivers from Y234C screen.** A. Western blot to check the protein levels. Gene knock-out by CRISPR-Cas9 was observed in post-invasion sgRNA expressing Y234C cells. B. Multiple DNA bands obtained from T7 endonuclease I assay indicating gene editing by CRISPR-Cas9.

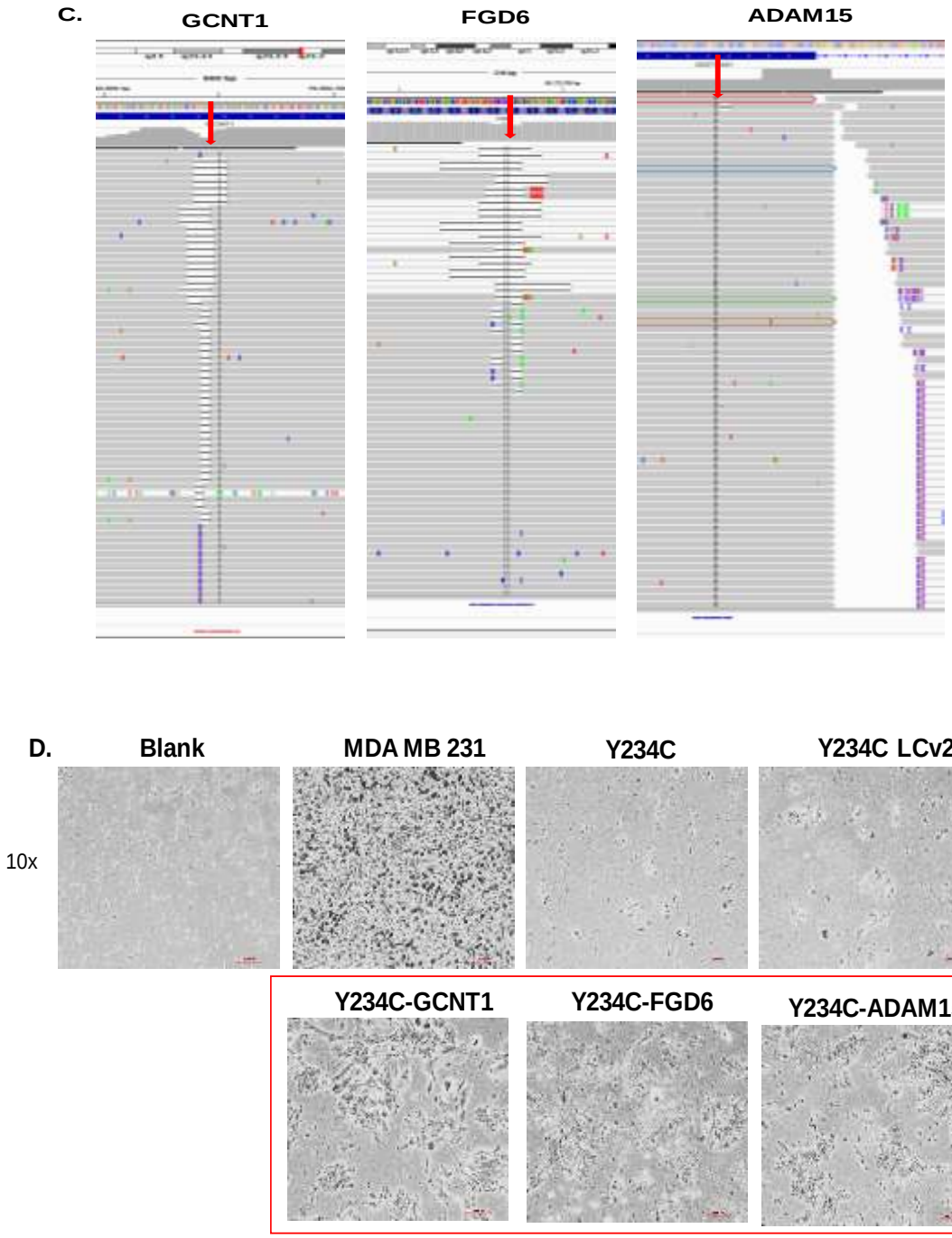
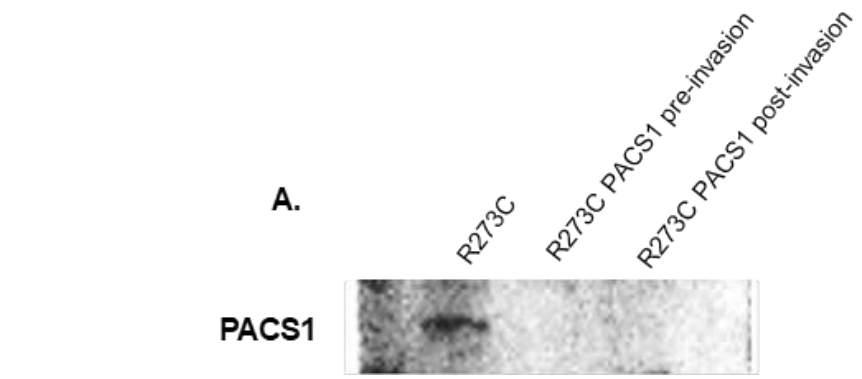


Figure 4.10. **Validation of selected co-drivers from Y234C screen.** C. NGS of the targeted gene loci show deletion of base pairs. At ADAM15 loci deletion could be at a distant target site. D. Disruption of GCNT1, FGD6 and ADAM15 increased the invasive capacity of weakly invasive Y234C cells. MDA-MB-231 cells were the positive control and Y234C LCv2 were the empty vector control.

Increased invasiveness of the isolated true CRISPR-Cas9 edited cells were confirmed with transwell invasion assay incubated for 24 hrs. The knock-out of *GCNT1*, *FGD6* and *ADAM15* in p53 mutant Y234C cells demonstrated increased invasion validating the results obtained from the invasion screening of the Y234C-libA and Y234C-libB cells (Fig 4.10D). The MDA-MB-231 cells and Y234C cells transduced with empty lentiCRISPRv2 plasmid with no sgRNA insert (Y234C LCv2) were used as the positive and negative controls, respectively. The Y234C LCv2 cells were not invasive, showing that constitutive expression of the Cas9 enzyme does not affect the invasion of the cells. These results demonstrated that the loss-of-function mutation of *GCNT1*, *FGD6,* and *ADAM15* drive increased invasion of mammary epithelial cells expressing p53-Y234C mutant protein.

Similarly, the R273C cell lines expressing the highly enriched sgRNAs targeting *PACS1* and *SDF4* obtained from post-invasion cell pool of R273C-libA and libB were also generated, and the invasive population was selected and tested for gene deletion, protein levels, and invasiveness. A significant decrease in the level of PACS1 protein was observed in both the pre and post-invasion fraction of R273C-PACS1 cells compared to R273C cells (Fig 4.11A). Several commercial antibodies for SDF4 were tried but did not work for Western blots. We also confirmed the deletion at the genomic loci of *PACS1* and *SDF4* by T7 endonuclease I assay (Fig 4.11B) and NGS (Fig 4.11C. Further, by using the same experimental setup for validating the top this for Y234C cells, we showed that the R273C LCv2 cells were slightly more invasive than the R273C cells but less than R273C sgRNA for *PACS1* and *SDF4* expressing cells. These individual validations showed that p53 missense mutant R273C collaborates with loss-of-function mutations in *PACS1* and *SDF4* to further drive cell invasion of mammary epithelial cells.



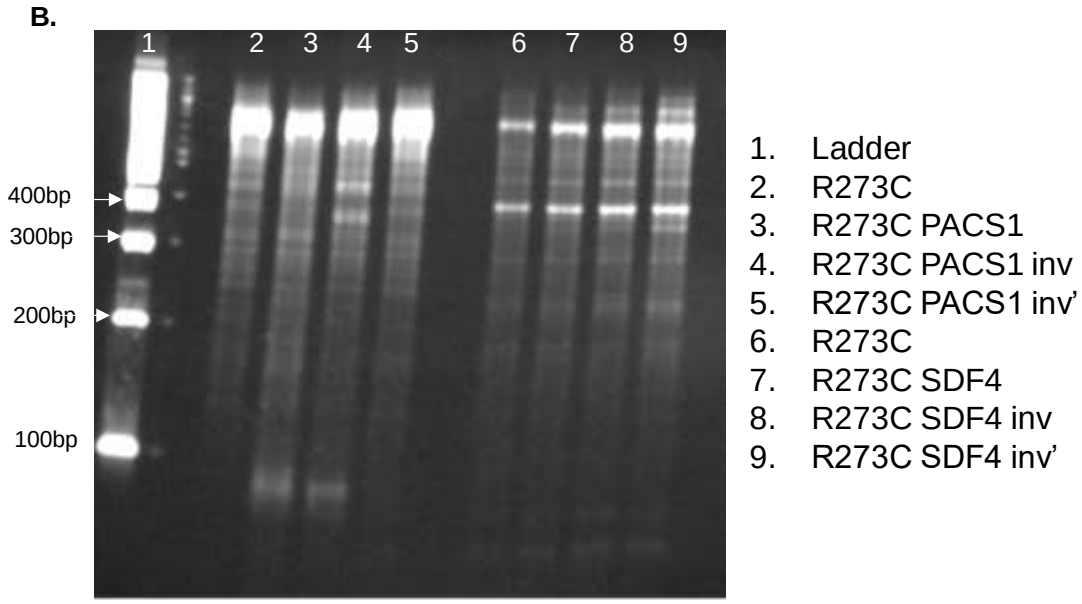


Figure 4.11. **Validation of selected co-drivers from R273C screen.** A. Western blot to check the protein levels. Gene knock-out by CRISPR-Cas9 was observed in post-invasion PACS1 sgRNA expressing R273C cells. Antibody for SDF4 did not work. B. Multiple DNA bands obtained from T7 endonuclease I assay indicating gene editing by CRISPR-Cas9.

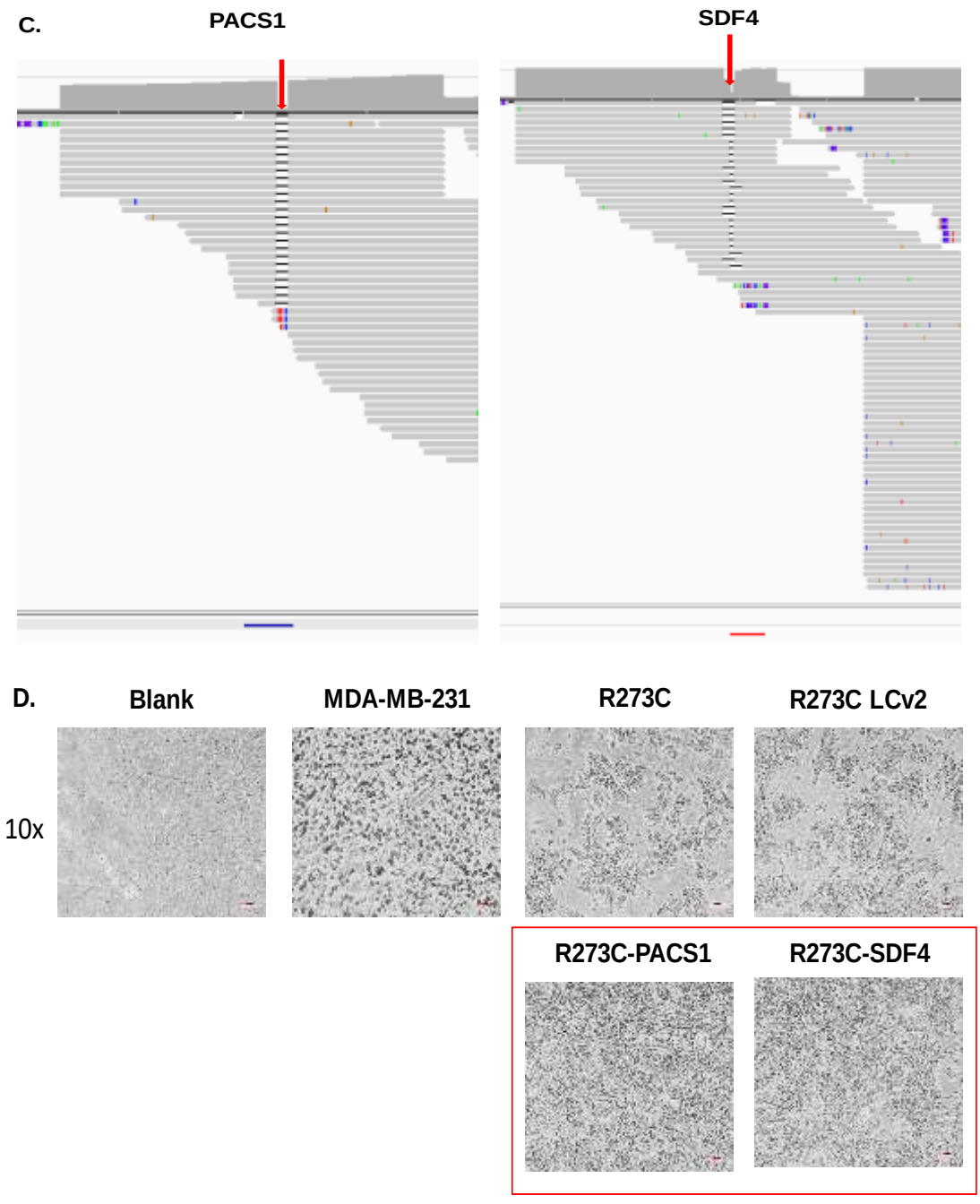


Figure 4.11. **Validation of selected co-drivers from R273C screen.** C. NGS of the targeted gene loci show deletion of base pairs. D. Disruption of PACS1 and SDF4 increased the invasive capacity of R273C cells. MDA-MB-231 cells were the positive control and R273C LCv2 were the empty vector control.

Conclusion

We established an efficient functional genomics pipeline to identify unique co-drivers of missense mutant p53 through a genome wide screen. Using this approach, we demonstrated that each missense mutant p53 cooperates with different co-drivers to enhance invasive cell phenotype.

The CRISPR-Cas9 based GeCKO v2 library provided an effective method to perturb every protein coding gene in humans for identification of possible co-driver mutations. We carefully designed the screening condition to target a 250X coverage for the library during screening for the co-drivers to ensure authenticity of the data. A well-defined method of isolating the post-invasion cell population and determining the highly enriched sgRNAs targeting specific genes in the post-invasion cell pool compared to pre-invasion or reference pool provided a competent way to capture the functionally relevant co-drivers for each p53 missense mutation. We used two MCF 10A cell-lines expressing p53 mutants Y234C and R273C with different invasiveness for the co-driver screen, and p53 WT-OE cells were included as a control. Screening of Y234C and R273C cells yielded different sets of enriched sgRNAs in the post-invasion cell pool that also differed in the order of magnitude of enrichment while no significant enrichment of sgRNAs was observed in the control p53 WT-OE post-invasion cell pool. These results further substantiated the fact that different set of mutant drivers and co-drivers collaborate in driving cancer tumor progression.

The top hits from the co-driver screen were carefully selected based on the number of times the candidate sgRNAs reappeared in independent screens, the targeting site of the sgRNA in the gene, any mutation co-occurrence with *TP53* mutations in patients and functional relevance to cell invasion, migration and cytoskeletal organization. Integrated analysis yielded *GCNTI*, *FGD6* and *ADAM15* as unique co-drivers of p53 Y234C while *PACS1* and *SDF4* were the co-driver candidates of p53 R273C enhancing the invasive potential of the cells. Individual validation of the co-driver candidates by generating stable cell lines with sgRNAs verified the results obtained from the genome-wide screen for increased cell invasion.

Our screening method captured the unique functionally relevant co-drivers of individual p53 missense mutants and explained how different combinations of drivers and co-driver mutations can contribute to inter or intratumor heterogeneity in cancer. Since we utilized the deletion-inducing sgRNA/Cas9 library in the current study, we could screen only for tumor suppressor-like co-drivers. However, our pipeline can be readily applicable for searching oncogene-like co-drivers just by using gene expression-activating libraries, for example.

We have observed that different p53 missense mutations produce differential phenotypes and can produce further functional and phenotypic alterations in the presence of other co-driver mutations. Our results from the co-driver screen show enrichment of different sets of co-drivers for two different p53 mutants highlighting the heterogeneity brought in by such unique combinations of mutations. Such a study can pave the way towards understanding of molecular heterogeneity observed in TNBC where *TP53* missense mutations are highly prevalent, and the information on co-driver combinations can be useful for designing alternative personalized targeted therapies for patients harboring a specific *TP53* mutation.

References

Brasacchio, Daniella, Amber E. Alsop, Tahereh Noori, Mariam Lufti, Sweta Iyer, Kaylene J. Simpson, Phillip I. Bird, Ruth M. Kluck, Ricky W. Johnstone, and Joseph A. Trapani. "Epigenetic control of mitochondrial cell death through PACS1-mediated regulation of BAX/BAK oligomerization." *Cell death and differentiation* 24, no. 6 (2017): 961.

Burdelski, Christoph, Michael Fitzner, Claudia Hube-Magg, Martina Kluth, Asmus Heumann, Ronald Simon, Till Krech et al. "Overexpression of the A Disintegrin and Metalloproteinase ADAM15 is linked to a Small but Highly Aggressive Subset of Prostate Cancers." *Neoplasia* 19, no. 4 (2017): 279-287.

Chen, Sidi, Neville E. Sanjana, Kaijie Zheng, Ophir Shalem, Kyungheon Lee, Xi Shi, David A. Scott et al. "Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis." *Cell* 160, no. 6 (2015): 1246-1260.

Chow, Ryan D., and Sidi Chen. "Cancer CRISPR screens in vivo." *Trends in cancer* 4, no. 5 (2018): 349-358.

Crevenna, Alvaro H., Birgit Blank, Andreas Maiser, Derya Emin, Jens Prescher, Gisela Beck, Christine Kienzle et al. "Secretory cargo sorting by Ca2+-dependent Cab45 oligomerization at the trans-Golgi network." *J Cell Biol* 213, no. 3 (2016): 305-314.

Gilbert, Luke A., Max A. Horlbeck, Britt Adamson, Jacqueline E. Villalta, Yuwen Chen, Evan H. Whitehead, Carla Guimaraes et al. "Genome-scale CRISPR-mediated control of gene repression and activation." *Cell* 159, no. 3 (2014): 647-661.

Hagisawa, Shigeru, Chikara Ohyama, Toshiko Takahashi, Mareyuki Endoh, Takuya Moriya, Jun Nakayama, Yoichi Arai, and Minoru Fukuda. "Expression of core 2 β1, 6-N-acetylglucosaminyltransferase facilitates prostate cancer progression." *Glycobiology* 15, no. 10 (2005): 1016-1024.

Hiles, Guadalupe Lorenzatti, Amanda Bucheit, John R. Rubin, Alexandra Hayward, Angelica L. Cates, Kathleen C. Day, Layla El-Sawy et al. "ADAM15 is functionally associated with the metastatic progression of human bladder cancer." *PloS one* 11, no. 3 (2016): e0150138.

Kang, Hua, Astrid Escudero-Esparza, Anthony Douglas-Jones, Robert E. Mansel, and Wen G. Jiang. "Transcript analyses of stromal cell derived factors (SDFs): SDF-2, SDF-4 and SDF-5 reveal a different pattern of expression and prognostic association in human breast cancer." *International journal of oncology* 35, no. 1 (2009): 205-211.

Korkmaz, Gozde, Rui Lopes, Alejandro P. Ugalde, Ekaterina Nevedomskaya, Ruiqi Han, Ksenia Myacheva, Wilbert Zwart, Ran Elkon, and Reuven Agami. "Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9." *Nature biotechnology* 34, no. 2 (2016): 192.

Miyamoto, Tsutomu, Akihisa Suzuki, Ryoichi Asaka, Kaori Ishikawa, Yasushi Yamada, Hisanori Kobara, Jun Nakayama, and Tanri Shiozawa. "Immunohistochemical expression of core 2 β1, 6-N-acetylglucosaminyl transferase 1 (C 2 G n T 1) in endometrioid-type endometrial carcinoma: a novel potential prognostic factor." *Histopathology* 62, no. 7 (2013): 986-993.

Platt, Randall J., Sidi Chen, Yang Zhou, Michael J. Yim, Lukasz Swiech, Hannah R. Kempton, James E. Dahlman et al. "CRISPR-Cas9 knockin mice for genome editing and cancer modeling." *Cell* 159, no. 2 (2014): 440-455.
Shalem, Ophir, Neville E. Sanjana, and Feng Zhang. "High-throughput functional genomics using CRISPR–Cas9." *Nature Reviews Genetics* 16, no. 5 (2015): 299.

Shalem, Ophir, Neville E. Sanjana, Ella Hartenian, Xi Shi, David A. Scott, Tarjei S. Mikkelsen, Dirk Heckl et al. "Genome-scale CRISPR-Cas9 knockout screening in human cells." *Science* 343, no. 6166 (2014): 84-87.

Solatycka, Alicja, Tomasz Owczarek, Friedrich Piller, Véronique Piller, Bartosz Pula, Lukasz Wojciech, Marzena Podhorska-Okolow, Piotr Dziegiel, and Maciej Ugorski. "MUC1 in human and murine mammary carcinoma cells decreases the expression of core 2 β1, 6-N-acetylglucosaminyltransferase and β-galactoside α2, 3-sialyltransferase." *Glycobiology* 22, no. 8 (2012): 1042-1054.

St Hill, Catherine A., Mariya Farooqui, Gregory Mitcheltree, H. Evin Gulbahce, Jose Jessurun, Qing Cao, and Bruce Walcheck. "The high affinity selectin glycan ligand C2-O-sLe x and mRNA transcripts of the core 2 β-1, 6-N-acetylglusaminyltransferase (C2GnT1) gene are highly expressed in human colorectal adenocarcinomas." *BMC cancer* 9, no. 1 (2009): 79.

Wan, Lei, Sean S. Molloy, Laurel Thomas, Gseping Liu, Yang Xiang, Sheree Lynn Rybak, and Gary Thomas. "PACS-1 defines a novel gene family of cytosolic sorting proteins required for trans-Golgi network localization." *Cell* 94, no. 2 (1998): 205-216.

Zhu, Shiyou, Wei Li, Jingze Liu, Chen-Hao Chen, Qi Liao, Ping Xu, Han Xu et al. "Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library." *Nature biotechnology* 34, no. 12 (2016): 1279.

CHAPTER 5

MOUSE MODEL-BASED SCREENING OF CO-DRIVERS OF MUTANT P53 FOR TUMOR GROWTH AND METASTASIS

Introduction

Cancer progression *in-vivo* is a multi-step clonal evolution and requires multiple mutations for tumor growth, invasion and metastasis. After establishing the *in-vitro* pipeline for identification of the co-drivers of mutant p53 that drive cell invasion (chapter 4), we defined an *in-vivo* mouse xenograft transplantation model to screen for the co-drivers of missense mutant p53 for tumor growth and subsequent metastasis.

The process of tumorigenesis is a rather complex process. A succession of several independent steps is needed for the successful completion of the tumorigenesis process. Transfection with at least two oncogenes was required for the transformation of rat embryonic fibroblasts. Co-transfection with *ras* and *myc* or *ras* and the gene for polyoma large T-antigen made cells tumorigenic that showed growth in soft agar and formed subcutaneous tumors in nude mice in 3 weeks (Land 1983). Similarly, mutation in p53 alone is not enough to drive tumorigenesis. Combinations of mutant p53 or deletion of *TP53* with other cancer genes are required. The p53 cDNA derived from a murine B-cell lymphoma (mutant p53) transfected alone in Rat-1 fibroblast cell-line could not induce foci formation in 2D cell cultures. But co-transfection with a *ras* oncogene transformed the cell readily. Injection of the Ras and mutant p53 expressing rat fibroblast cells formed aggressively growing tumors in the nude mice (Parada 1984). Transgenic mice with p53 loss and expression of specific oncoproteins such as *c-Myc* and *Ras* or with knock-out of tumor suppressors like *Rb, Nf1 or BRCA1* show accelerated rate of tumor development compared to the individual genetic lesions. These double or compound mutants have shortened latency, increased penetrance of tumorigenesis and develop novel tumor types (Attardi 2005).

In contrast to the primary rodent cells, normal human epithelial and fibroblast cells required additional expression of *hTERT* (telomerase catalytic subunit) that immortalized the cells and two oncogenes SV40 large T oncoprotein and *H-ras* for tumorigenic conversion (Hahn 1999). The primary human mammary epithelial cells (HMECs) with the same combination of genetic elements formed poorly differentiated subcutaneous or mammary carcinomas in mice with average latency of 7.5 weeks. The transformed primary HMECs when mixed with matrigel or primary human mammary fibroblasts increased tumor formation and decreased latency period highlighting the importance of stromal microenvironment on tumor formation (Elenbaas 2000).

c-Myc is more frequently amplified in ER-, PR- or the basal subtype of breast cancer that also harbors frequent mutations in *TP53* and higher expression of MYC correlates with poorer clinical outcome (Liao 2000).The basal like or the tripe negative breast tumors have very high mutation rates and harbor more mutations than the luminal subtype tumors. HER2, EGFR and MYC are commonly amplified oncogenes in breast cancer and co-amplification of MYC/HER2 is associated with high grade and shortened survival (Al-Kuraya 2004). For most of the basal tumors, the partnering mutated genes other than the commonly known EGFR and MYC, of p53 mutants remain elusive. The TNBC cell-lines that harbor *TP53* mutations (deletion or missense mutations) also overexpress EGFR or MYC (Chavez 2010) for example the MDA-MB-468, MDA-MB-436 and HCC1395 cell-lines. On the other hand, in patient derived tumor samples there could be other unknown less prevalent partnering genes of mutant p53 driving tumor development.

Aim

To test if p53 mutation with co-driver combinations are sufficient to support tumor growth *in-vivo* and perform *in-vivo* screens for identification of additional co-drivers.

The mutations in *TP53* alone are neither enough to promote transformation of cells nor tumor formation. Deletion of *PTEN* in p53 mutant expressing cells form colonies in soft agar demonstrating the transformation of the cells (Kim 2015). Many somatic gene mutations co-occur with mutation in *TP53* in cancer tumors and collaborate with mutant p53 to drive tumor initiation and subsequent metastasis but which combination of co-mutations or the co-drivers partner with mutant p53 largely remain unknown.

Approach

Xenograft mouse model and method for cell injection

Breast cancer tumor growth and metastasis can be modeled in different ways. For the identification of co-drivers of mutant p53 that support tumor growth, we chose the cell line xenograft transplantation mouse model. Implantation of the cell lines orthotopically (into mammary fat pads of mice) or ecto-topically (subcutaneously) into immunocompromised mice can form primary tumors followed by metastasis resembling multiple stages of malignant breast cancer development in patients. The immunodeficient mice such as NOD/SCID, NOG or NSG show high rate of cell engraftment and lead to tumor growth and metastasis (Ito 2012).

Injection of cells into the mammary gland provide a favorable microenvironment to model breast cancer development. This method is simple, less time consuming than developing genetically altered mice and allows implantation of multiple gene manipulated cell lines. The xenograft model can rapidly analyze the tumorigenic and metastatic potential of experimental cells, response to a therapeutic regime and resembles the heterogeneity of human tumor cells.

Testing invasion promoting co-drivers of mutant p53 for tumor growth in-vivo

We speculated that MCF 10A cells expressing mutant p53 with *PTEN* deletion that formed colonies in soft agar demonstrating cell transformation could form tumors in mice. We hypothesized that the post-invasion cell population from the CRISPR based screening of co-drivers of mutant p53 (with different combination of co-drivers and mutant p53, described in chapter 4) could form tumors and may metastasize as the cells were already proliferative and invasive. If these injected cells formed tumors in mice, the parallel sequencing of the primary tumors and the metastasized tumors would reveal the specific set of gene mutations collaborating with mutant p53 and responsible for the metastatic spread of tumors in the given cellular context.

For this *in-vivo* study, we required an appropriate xenograft mouse model and first started with optimization of the conditions needed to establish such a model. Once the parameters were set, different experimental cell lines (p53 mutation with co-drivers) were injected into the mice and were observed for the appearance of tumors.

Addition of oncogenes for the in-vivo screens

We used the GeCKO v2 CRISPR based library, a gene knock-out library, to identify the possible co-drivers of mutant p53 driving cell invasion. The resultant mutated co-drivers obtained from the cell invasion screen yielded only the tumor suppressor genes as the possible co-drivers in the context of different p53 missense mutations. The tumor cells harbor both mutated tumor suppressors and oncogenes. Our experimental cell lines as described above have a combination of at least two mutated tumor suppressor genes (mutant p53 + co-driver gene knock-out) which may not form tumors in the immune-compromised NOD/SCID mice. The TNBC cell lines known to be tumorigenic and metastatic, harboring a mutation in *TP53* or *TP53* and *PTEN*, contain amplification of either *MYC* or *EGFR* that are the oncogenes (for example, MDA-MB-468 and HCC1395). If the experimental cell-lines expressing mutant p53 with co-driver combinations could

not form tumors, over-expression of an oncogene EGFR or MYC in the cells may induce tumor growth.

Method

Maintaining mice breeding colonies

To conduct all the mice experiments, maintenance of breeding colonies was essential. The NOD/SCID mice were bred at 2 or 3 months of age as they get non-productive by the age of 6 months. The number of mice bred were dependent on the total number needed for an experiment at a given time. The approximate litter size for NOD/SCID mice is 8 to 10 and so, the breeding pairs were carefully set to get the required number of females for the experiments and extra to retain as breeders to maintain the line. Breeding was done in compliance with the ASU Mouse Breeding Cage Density Guideline. The male was removed upon confirmation of pregnancy and the pups were weaned by 21 days.

Cell injection methods

For each experiment the cells were harvested by trypsin treatment (0.25% trypsin) and then suspended in media with Matrigel (1:1) and kept on ice throughout the following procedures. The NOD/SCID female mice were injected with the appropriate concentration of cells either subcutaneously on the flank, directly in the mammary fat pad or through the tail vein.

Cell injection directly into mammary fat pads

We chose to proceed with our experiments using the mammary fat pad injections. The mice surgeries were carried out by a certified veterinarian from Department of animal care and

technologies, ASU. To establish the xenograft mouse breast cancer model, 5 to 8 weeks old female NOD/SCID mice were injected with the experimental cell lines. Mice have 10 mammary glands, 5 on each side, located ventro-laterally. They can be distinguished as the thoracic glands, 1 to 3, and inguinal glands, 4 and 5. The inguinal gland 4 is most easily accessible and hence, commonly used for clearing and transplantation experiments. Before cells were transplanted, it was necessary to completely remove the endogenous epithelial tissue from the mammary pad. Each mouse was anesthetized with ketamine, laid on one side and the skin were shaved and prepared aseptically. A 5 mm incision was made by small scissors around the 4^{th} nipple. The skin and the nipple were removed together with a region containing the closest part of the fat pad and the mammary anlage. A pocket under the skin was opened in a cranial direction with the blades of the scissor so that the mammary fat pad was seen. 100 ul of suspended cells in Matrigel were injected in the fat pads within 40 min of harvesting. It was made sure that the inoculum remained inside the fat pad and did not leak into the subcutaneous space. The incision was closed with suture (4-0 nylon) or surgical glue and each mouse was monitored and kept warm while recovering from anesthesia.

Imaging of mice and end point criteria

NOD/SCID mice are immune compromised and were kept in sterile cages. Each mouse was marked with picric acid to determine the cell lines they were injected with. To determine the timeline of tumor onset and metastasis, *in-vivo* imaging of the cells was performed using the RGD-IR dye (LI-COR 926-09889) that binds to the integrin receptor (specific to transformed cells) in the XPearl Impulse Imaging System routinely used at ASU. The IR (infrared) dye has low non-specific binding to cellular components and high signal-to-noise ratio. This *in-vivo* imaging system is specifically designed for small animal imaging and utilizes near infrared (NIR) probes for imaging which have very low background for auto-fluorescence in whole animals. Hence, this imaging system allows sensitive detection of NIR labeled antigens/ adjuvants injected into the animals with good signal to noise ratio. The mice were shaved with electric clippers over the

injection area (1.5 cm X 1.5 cm) to image the tumor by reducing the background effect caused by animal fur. The mice were always anesthetized with isoflurane at 1 to 2.5% inhalation for the Retro-Orbital (RO) injection (following the IACUC SIG for Retro-Orbital injection of mice) with 30ul dye during imaging. Once primary tumors appeared, the mice were monitored every other day. Once the primary tumor size measured 1.5 cm in diameter, tumors were measured daily, and mice were euthanized before the tumors measured 2 cm in diameter. Metastasis at this stage was tracked through weekly dye injections and imaging. Mice with metastasis often showed hunched posture, ruffled fur, non-mobility, labored respiration (dyspnea) and loss of body weight. These mice were immediately euthanized and the primary tumor as well as the organs with metastasized cells were kept in formalin solution. Mice were euthanized with CO_2 inhalation followed by cervical dislocation at the end of each experiment or when required.

Results

The pilot project to establish the xenograft mouse model

In a pilot project we tried to optimize the parameters to establish a xenograft mouse model to study the tumorigenic and metastatic (if tumorigenic) potential of cell lines bearing different combinations of mutations (p53 mutation and co-drivers) using the immunocompromised female NOD/SCID mice. These mice have impaired B and T cell lymphocyte development and deficient in natural killer cells. This leads to the high acceptance of the engraftments (cells or tissues from other species) in these mice. We optimized the injection route, cell concentration injected into the mice, volume of cells per injection and mice imaging technique using the metastatic TNBC cell line MDA-MB-231 as the control. The cells harbor missense mutation in *TP53* along with numerous other somatic mutations and form tumors in NOD/SCID mice that metastasize to the lungs and liver. The results from the pilot experiment is shown in the Table 5.1.

Table 5.1. **Pilot project to establish the xenograft mouse model.**

Cell-line	No. of NOD/SCID female mice	Cell conc.	Injection site	Tumor appearance	Metastasis
MDA-MB-231 (positive control)	2	200,000	Subcutaneous flank	Yes	Lung, liver
MDA-MB-231	2	$2X10^6$	Mammary fat pad	Yes	Lung, liver
MDA-MB-231	3	$2X10^6$	Tail vein	Yes	All over the body

Visible and measurable tumors in the flank and the mammary fat pad were observed for the control TNBC cell line MDA-MB-231 after 4 weeks following cell injection. Metastasis was observed between 6 to 10 weeks as the health of the mice rapidly deteriorated (loss of body weight and immobility). Tumors with subcutaneous injection in the flank with a million cells grew very rapidly and hence, the number of cells were reduced to 200,000 for the highly proliferative MDA-MB-231 cell line to allow the monitoring of metastasis for up to 10 weeks. The tail vein injections resulted in the spread of tumors across the body and the mice died by 6 weeks after cell injection. Tumors from the primary site (flank or mammary fat pads) spread mainly to the lungs and the liver, the usual sites for breast cancer metastasis in mice.

We tested the tumorigenicity of most of our experimental cell-lines through subcutaneous flank injections as they are easy to perform and tumors (when and if they appear) are easy to monitor. Though the primary tumors and metastasis appeared for MDA-MB-231 cells through both the subcutaneous and mammary fat pad injections, we chose to use the orthotopic mammary fat pad injections for our experimental cell lines as it is more relevant for studying breast cancer tumor formation and metastasis. This xenograft mouse model was used to test the tumorigenic ability and possible metastasis of our experimental cell lines with combination of p53 mutation and co-drivers and at least 2 mice were used per condition in each experiment as replicates.

Testing the growth of the cells in-vivo with defined co-driver of p53 mutation (invasive MCF 10A p53 mutant cell-lines and 10A p53 mutant PTEN^{null} cells)

We first started the *in-vivo* experiments by injecting invasive MCF 10A p53 mutant R248W, R273C and Y220C cell lines that were found to be the most aggressive p53 mutants determined from the *in-vitro* cell assays (described in chapter 2, page 93). The cells were injected subcutaneously in the flank of the female NOD/SCID mice. None of the cell lines formed tumors in mice by 14 weeks after the injections. Such a result was expected as *TP53* mutations alone cannot transform cells *in-vitro* and form tumors *in-vivo*.

In the next set of experiments, the MCF 10A p53 mutant Y234C cells with *PTEN* deletion (*PTEN* knocked out using CRISPR-Cas9 system as known co-driver of mutant p53) with high invasion phenotype were injected subcutaneously (Y234C PTEN-C4 and Y234C PTEN-C5) and observed for 10 weeks. The p53 mutant Y234C cells were very weakly invasive but the Y234C cells also harboring a deletion mutation in *PTEN* could invade the matrigel barrier efficiently and showed anchorage independent growth in soft agar. But these cells did not form any tumors in mice. We expected the cells to be tumorigenic as MCF 10A *TP53*^{null} *PTEN*^{null} cells form tumors and metastasize in a NOD/SCID mouse xenograft model. The tumors appeared in mice by 4 weeks and formed huge primary tumors and metastasized to lungs by 8 weeks after injection into the mammary fat pads as reported by Kim et.al, (2015). No tumors formed by the Y234C-PTEN null cells by week 13. The MDA-MB-231 control cells formed tumors by the 4th week and the mice were euthanized by week 8 due to metastasis of the cells to the lungs (Fig 5.1, Table 5.2).

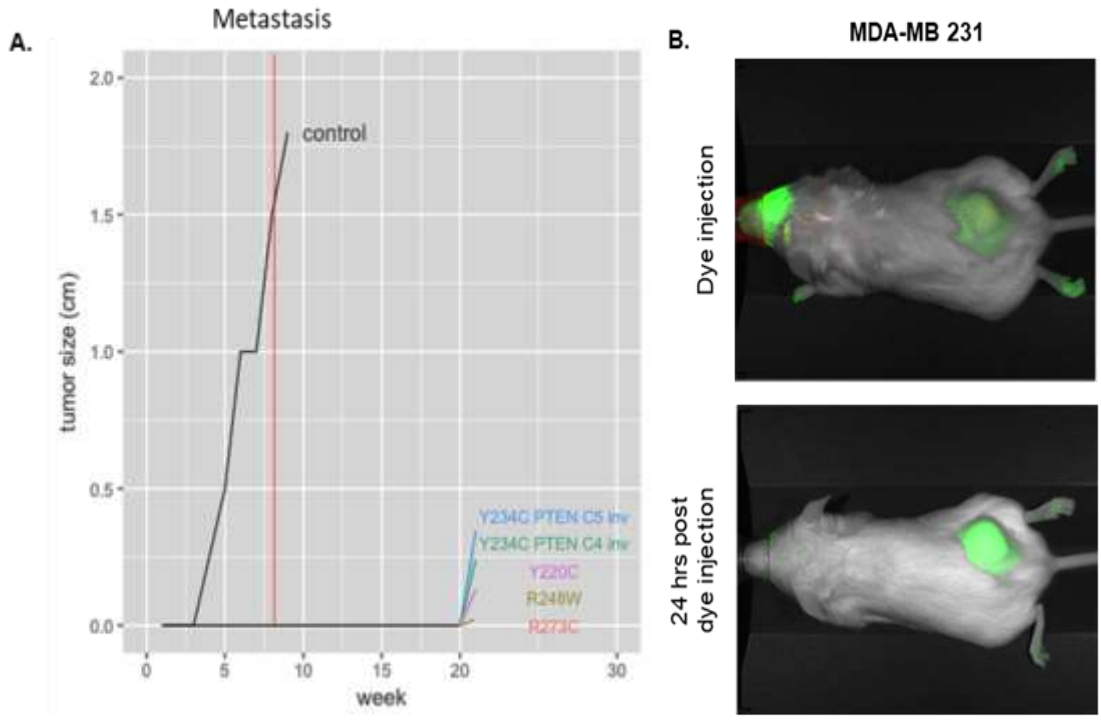


Figure 5.1. **Results for mice injections with aggressive MCF 10A p53 mutants and mutant p53 *PTEN*^{null} cells. A.** The highly aggressive p53 mutants and Y234C PTEN C4 and Y234C PTEN C5 (with sgRNAs 4 and sgRNA 5 targeting *PTEN* gene) were not tumorigenic. MDA-MB-231 cells were positive controls and formed tumors by week 4. By 8 weeks cells metastasized (red line) and the mice were euthanized. B. Localization of the Li-Cor dye at the growing tumor site (subcutaneous, flank) seen in mice injected with MDA-MB-231 cells 24 hrs post dye injection.

Table 5.2. **Mice injection detail with the aggressive MCF 10A p53 mutants and MCF 10A cells expressing mutant p53 with deletion in *PTEN*.**

Cell-line	No. of NOD/SCID female mice	Cell conc.	Cell type	Injection site	Time for exp	Tumor appearance	Metastasis
MDA-MB-231	2	$2X10^6$	Positive control	Subcutaneous flank	6 weeks	Yes	Lung, liver
R248W	2	$3X10^6$	Experimental	Subcutaneous flank	14 weeks	No	
R273C	2	$3X10^6$	Experimental	Subcutaneous flank	14 weeks	No	
Y220C	2	$3X10^6$	Experimental	Subcutaneous flank	14 weeks	No	
Y234C PTEN C4 inv	2	$1X10^6$	Experimental	Subcutaneous flank	10 weeks	No	
Y234C PTEN C5 inv	2	$1X10^6$	Experimental	Subcutaneous flank	8 weeks	No	

In-vivo screen with CRISPR library transduced MCF 10A p53 mutant cells transduced at LMOI

We set up a genome wide screen for the identification of biologically important but less frequent driver mutations, the co-drivers, cooperating with the recurrent *TP53* mutations in the context of mammary epithelial cells for increased cell invasion (described in chapter 4). For the screening, MCF 10A p53 mutants Y234C and R273C cells were transduced with the GeCKO v2 library (CRISPR-Cas9 based with sgRNAs targeting every protein coding gene in humans) at low multiplicity of infection (LMOI) and used for cell invasion assay to identify the co-drivers promoting invasive phenotype. As control for the screening, we included the MCF 10A p53 WT-OE cells that over-expressed wildtype p53.

Only the invasive cells that managed to cut through the Matrigel barrier were isolated. The post-invasion cell population showed enrichment of different sgRNAs compared to the pre-invasion cell pool. We hypothesized that the post-invasion cell pool with

different co-driver combinations (*TP53* mutation with an additional mutation) could form tumors when injected into mice. We injected the isolated post-invasion CRISPR library transduced cell population (transduced at low MOI) in the mice for the next set of experiments.

The post-invasion (invasive) library transduced cells (WT-OE libA, WT-OE libB, Y234C libA and Y234C libB) cells were injected in the flank of female NOD/SCID mice subcutaneously and monitored for tumor formation. These cells did not form any tumors by week 20. MDA-MB-231 positive control cells formed tumors and metastasis by 8 weeks as observed in the previous experiments (Fig 5.2, Table 5.3).

Table 5.3. **Mice injections with MCF 10A cells transduced with GeCKO v2 library at low MOI.**

Cell-line	No. of NOD/SCID female mice	Cell conc.	Cell type	Injection site	Time for exp	Tumor appearance	Metastasis
MDA-MB-231	2	200,000	Positive control	Subcutaneous flank	8 weeks	Yes	Lung, liver
WT-OE libA inv	2	$3.5X10^6$	Experimental	Subcutaneous flank	20 weeks	No	
WT-OE libB inv	2	$3.5X10^6$	Experimental	Subcutaneous flank	20 weeks	No	
Y234C libA inv	2	$3.5X10^6$	Experimental	Subcutaneous flank	20 weeks	No	
Y234C libB inv	2	$3.5X10^6$	Experimental	Subcutaneous flank	20 weeks	No	

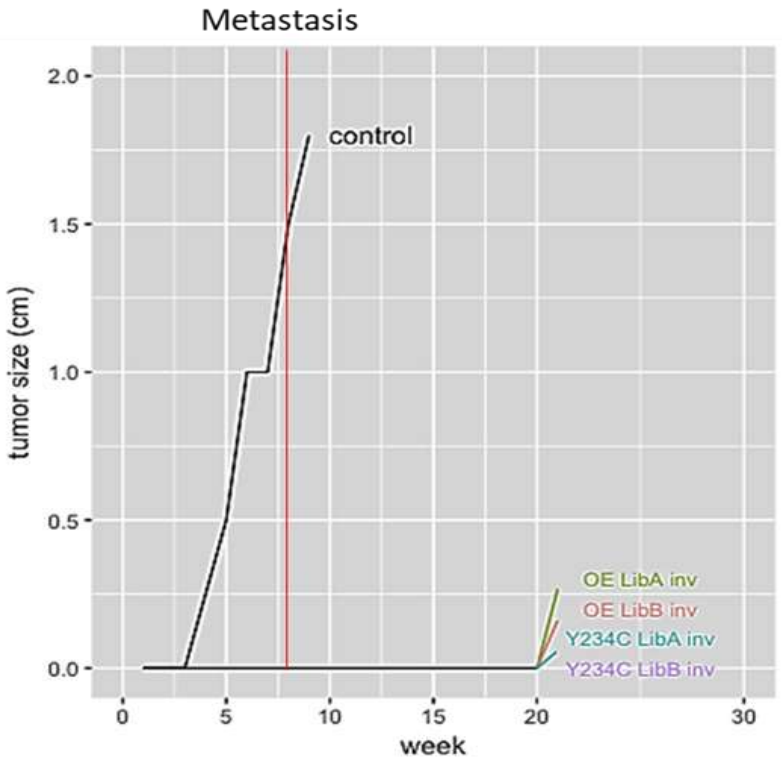


Figure 5.2. **Mice injections with MCF 10A p53 mutant cells transduced with GeCKO v2 library at low MOI.** None of the cells were tumorigenic. MDA-MB-231 cells were positive controls and formed tumors by 4 weeks. By the week 8 cells metastasized (red line) and the mice were eventually euthanized.

In-vivo screen with CRISPR library transduced MCF 10A p53 mutant cells transduced at HMOI

From the previous results we observed that the post-invasion MCF 10A Y234C library transduced cells (at low MOI), Y234C libA and Y234C libB, could not form any tumors in mice. A combination of two mutations (p53 mutation and a co-driver) were not enough to drive tumor formation. Moreover, Y234C is a weakly invasive p53 missense mutant. We anticipated that transduction of the most aggressive MCF 10A p53 mutant

R273C cells (determined by cell-based assays, chapter 2) with GeCKO v2 library (low MOI) would result in different co-driver combinations with enhanced possibility of tumor formation *in-vivo*.

We injected the post-invasion R273C libA and R273C libB cell population (transduced at low MOI) in the mammary fat pads of the female NOD/SCID mice. We speculated that with an aggressive p53 mutation, R273C, the new mutation combinations would generate tumors compared to less aggressive Y234C p53 mutant. But no tumors were observed by week 16 since the injections (Y234C library transduced cells were observed for 20 weeks).

As the library transduced cells at LMOI (only one additional sgRNA in cells with p53 mutation) failed to form any tumors, it was clear that only two mutations were not enough to turn cells tumorigenic in the given context. Hence, we decided to transduce the highly aggressive MCF 10A p53 mutant R273C cells with the GeCKO v2 (CRISPR-Cas9 based) library at high multiplicity of infection conditions (HMOI, MOI > 4). This ensured that the transduced cells acquired more than two sgRNAs or mutations at a time in addition to the p53 mutation increasing the probability of tumor formation. The library transduced (at high MOI or HMOI) R273C libA (HMOI) and R273C libB (HMOI) cell populations both pre-invasion and post-invasion were orthotopically injected into the mammary fat pad of the female NOD/SCID mice. But even with a combination of at least three or more than three mutations, no tumors were seen after 20 weeks of cell injection. R273C cells were used as the negative controls (they do not form tumors) and the MDA-MB-231 cells as the positive that formed metastasis by 8 to 9 weeks (Fig 5.3, Table 5.4).

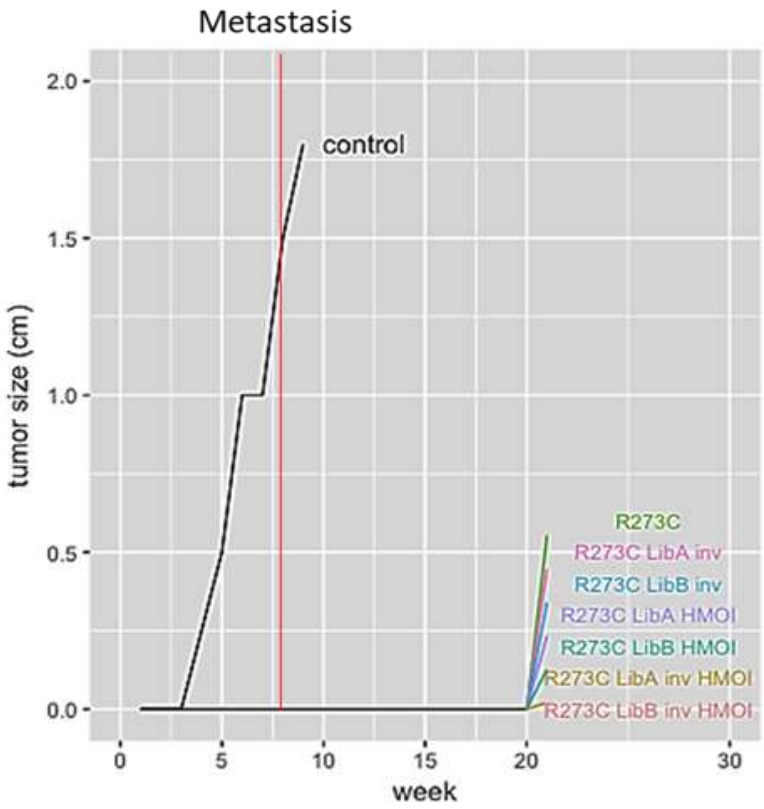


Figure 5.3. **Mice injections with MCF 10A p53 mutant R273C transduced with GeCKO v2 library at low and high MOI (HMOI).** Inv refers to invasive or post-invasion cells. None of the cells were tumorigenic. MCF 10A R273C cells were the negative control and MDA-MB-231 cells the positive control and formed tumors by week 4. By the 8 weeks cells metastasized (red line) and the mice were euthanized by 10 weeks.

Introduction of an oncogene in the experimental MCF 10A mammary epithelial cells with combination of mutant tumor suppressor genes

The GeCKO v2 CRISPR based library is a gene knock-out library and the resultant mutant co-drivers obtained from the cell invasion screen yielded only the tumor suppressor genes as the possible co-drivers in the context of different p53 missense

mutations. But the tumor cells harbor both mutated tumor suppressors and oncogenes. Our experimental cell lines as described above have a combination of at least two

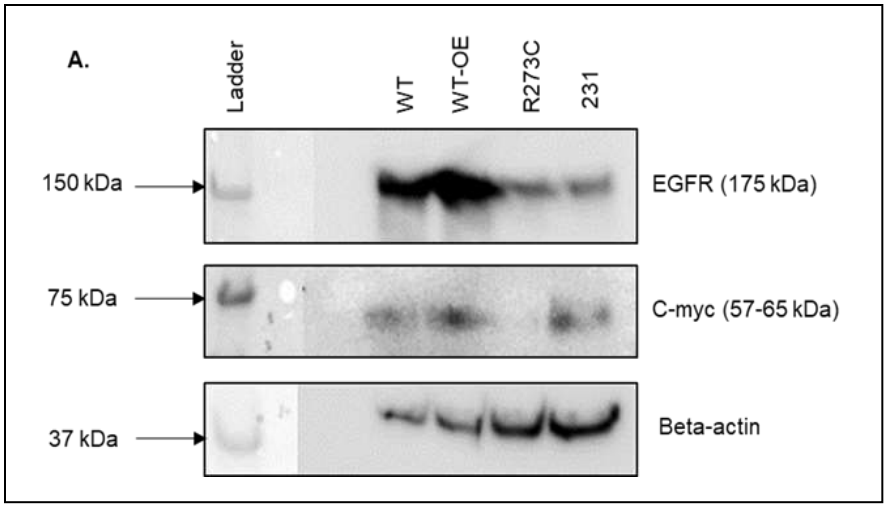
Table 5.4. **Mice injections with MCF 10A p53 mutant R273C cells transduced with GeCKO v2 library at low (LMOI) and high MOI (HMOI).**

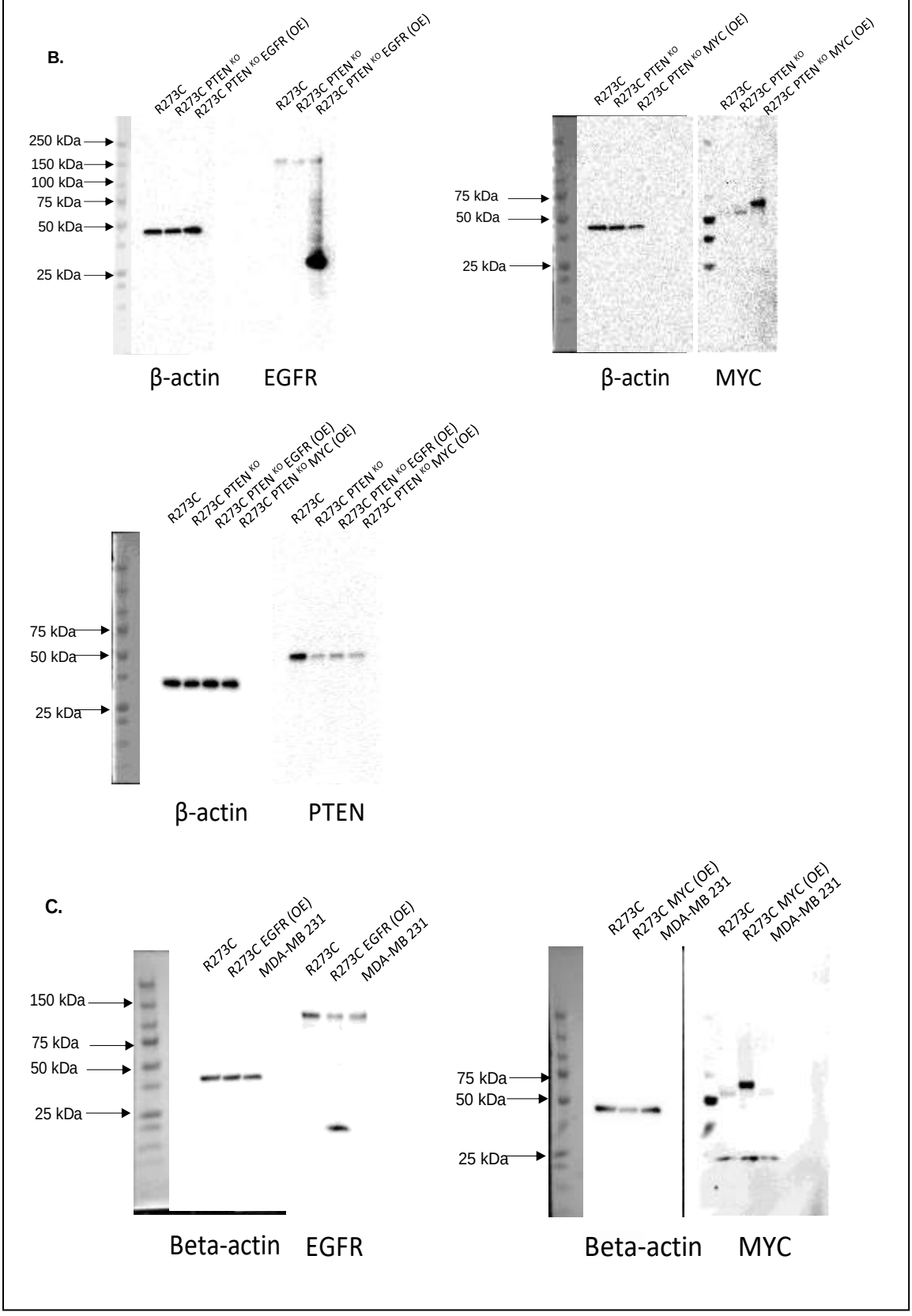
Cell-line	No. of NOD/SCID female mice	Cell conc.	Cell type	Injection site	Time for exp	Tumor appearance	Metastasis
MDA-MB-231	2	$1X10^6$	Positive control	Mammary fat pad	8 weeks	Yes	Lung, liver
R273C	2	$3X10^6$	Negative control	Mammary fat pad	16 weeks	No	
R273C libA inv	2	$3X10^6$	Experimental	Mammary fat pad	16 weeks	No	
R273C libB inv	2	$3X10^6$	Experimental	Mammary fat pad	16 weeks	No	
R273C LibA HMOI pre-inv	2	$1.25X10^6$	Experimental	Mammary fat pad	20 weeks	No	
R273C LibB HMOI pre-inv	2	$1.25X10^6$	Experimental	Mammary fat pad	20 weeks	No	
R273C LibA HMOI inv	2	$1.25X10^6$	Experimental	Mammary fat pad	20 weeks	No	
R273C LibB HMOI inv	2	$1.25X10^6$	Experimental	Mammary fat pad	20 weeks	No	

mutated tumor suppressor genes (due to the use of knock-out CRISPR library) which proved to be insufficient to form tumors in the immune-compromised NOD/SCID mice. The TNBC cell lines known to be tumorigenic and metastatic with a mutation in *TP53* or *TP53* and *PTEN* also contain amplification of the oncogene either *MYC* or *EGFR* (for example, MDA-MB-468 and HCC1395). We thought of introducing an oncogene, either MYC or EGFR, in our experimental cell lines to boost the probability of tumor formation.

We first tested the protein level of EGFR and MYC in the MCF 10A cells with endogenous wildtype p53, p53 WT-OE and R273C cells including the MDA-MB-231 control cells using western blot (Fig 5.4A). EGFR protein level in R273C cells and MDA-MB-231 cells were comparable. There was no detectable MYC in the R273C cells. We generated a set of cell lines (Table 5.5, with different combinations of tumor suppressors and oncogenes) over-expressing (OE) EGFR or MYC in addition to the p53 and *PTEN* mutations (Fig 5.4B). Another set of cell-lines included the R273C over-expressing EGFR or MYC followed by transduction with the GeCKO v2 libA and libB at high MOI (Fig 5.4C). The cells (with various mutation combinations) were then injected into the mammary fat pads of female NOD/SCID mice.

EGFR and MYC were over-expressed using mammalian expression vector pLenti6.3 that adds a V5 tag at the C-term end of the protein. The over-expressed EGFR was a truncated version while MYC was overexpressed with a V5 tag (Fig 5.5B). The mice transplanted with cells over-expressing MYC would likely form tumors than the EGFR expressing cells. The cells forming primary tumors and eventual metastasis can then be utilized to screen for the specific mutated gene combinations driving tissue-specific breast cancer metastasis.




B.
R273C
R273C PTEN KO
R273C PTEN KO EGFR (OE)
250 kDa
150 kDa
100 kDa
75 kDa
50 kDa
25 kDa
β-actin
EGFR
R273C PTEN KO MYC (OE)
MYC
R273C PTEN KO EGFR (OE) MYC (OE)
PTEN
C.
MDA-MB 231
EGFR (OE)
MYC (OE)
Beta-actin

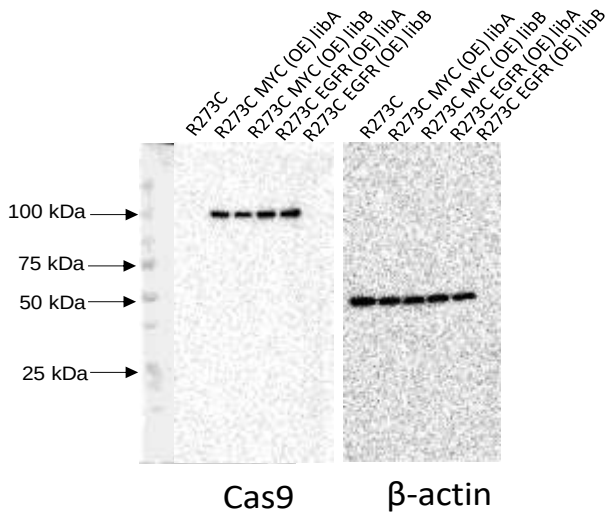


Figure 5.4. **Western blot images to check level of EGFR, MYC and Cas9.** A. Western blot to check levels of endogenous EGFR and MYC. B. Expression levels of EGFR, MYC and PTEN in the different mutation combination cells generated for mice injections. PTEN was first knocked out using sgRNA in p53 mutant R273C cells. Afterwards MYC or EGFR was over-expressed in the cells. C. R273C cells were first transduced to over-express MYC or EGFR and then transduced with the CRISPR based GeCKO v2 half libA and B. Only the library transduced cells expressed Cas9. Beta-actin was the loading control. OE over-expression, KO knock-out.

Table 5.5. **Cell-lines generated with over-expression of EGFR and MYC. OE over-expression, HMOI high multiplicity of infection and del deletion.**

Tumorigenic?	Cell conc. Injected
R273C + PTEN (del)	$2.8X10^6$
R273C + PTEN (del) + EGFR (OE)	$2.8X10^6$
R273C + PTEN (del) + MYC (OE)	$2.8X10^6$
R273C + EGFR (OE) + libA (HMOI)	$2.8X10^6$
R273C + MYC (OE) + libA (HMOI)	$2.8X10^6$
R273C + EGFR (OE) + libB (HMOI)	$2.8X10^6$
R273C + MYC (OE) + libB (HMOI)	$2.8X10^6$

Interestingly, one out of the two mice injected with R273C MYC libB and R273C EGFR libA cells formed huge primary tumors (1.5 cm in diameter) by week 9. The tumors were massive, and the mice were euthanized. During necropsy, no metastasis was observed in the mice and the primary tumors were collected. The primary tumors were cut into tiny sections and stored at -80^{O}C for subsequent processing for single cell sequencing. As the cells contain multiple mutations, bulk sequencing would not be appropriate to decipher the mutation combinations that formed the tumors. Single cell sequencing of at least 1000 cells (from different sections of the tumor) would reveal the mutations that generated tumors in the mice. We intend to repeat the experiment with the cells that formed the tumors in mice with a larger set (6 mice instead of 2) to confirm the results. The positive control cells MDA-MB-231 formed tumors by week 5 that metastasized by week 8. The results showed that combination of oncogenes and mutated tumor suppressors drive tumor formation.

Conclusion

From the CRISPR based screen for ‘co-drivers’ of missense mutant p53, we identified important co-driver gene candidates specific to different p53 mutant background that enhanced invasion potential of cells (described in chapter 4). Cell migration and invasion sets the initiating stages for the metastatic cascade. Once we extracted the post-invasion cell population, we wondered if these invasive cells could initiate tumor formation when transplanted into immune-compromised NOD/SCID mice. After establishing the parameters for a xenograft NOD/SCID mouse model for tumor formation and metastasis (cell concentration, injection route and timeline) we tested several experimental cell lines with different combinations of mutations for tumor growth. We used the metastatic TNBC cell line MDA-MB-231 as the positive control for

all our experiments that rapidly formed tumors by 4 to 5 weeks and metastasis by 8-10 weeks in the lungs and the liver in the NOD/SCID mice. These cells are very aggressive and so, we generated the MCF 10A cells harboring *PTEN* deletion and over-expression of EGFR or MYC in addition to p53 mutation as the new control cells. Combination of MYC and mutant p53 transform cells for anchorage independent growth in soft agar. MCF 10A cells with these mutation combinations would provide a better control for the *in-vivo* experiments.

The highly aggressive MCF 10A p53 mutant R273C, R248W and Y220C cells injected subcutaneously in the flank did not form any tumors in the NOD/SCID mice. This result was expected as mutation in p53 alone cannot form colonies in soft agar to show anchorage independent growth for cellular transformation. We generated cell lines with *PTEN* knock-out in weakly invasive p53 mutant Y234C that resulted in increased invasive potential of the cells. But no tumors were observed for the mice injected with Y234C *PTEN* ^{KO} cells. Such a result was unexpected as cells with downmodulation of both p53 and PTEN are tumorigenic when transplanted in mice within 8 to 10 weeks of injection. An explanation could be that *PTEN* deletion is frequent in TNBC subtype of breast cancer which is the most aggressive of all breast cancer subtypes and harbors either deletion mutation or other hotspot missense mutations in *TP53*. Y234C is not a hotspot *TP53* missense mutation and from our phenotypic assays based on the hallmarks of cancer (chapter 2), Y234C is a much less aggressive mutant. Hence, a combination of Y234C and *PTEN* knock-out might not be strong enough to turn cells tumorigenic though such double mutant cells formed colonies in the soft agar assay and demonstrated increased cell invasion compared to just Y234C mutation.

After the isolation of the post-invasion cell population of Y234C libA and Y234C libB transduced at low multiplicity of infection (LMOI), we transplanted the cells (p53 mutation with combination of different co-drivers) subcutaneously into the flank of the

NOD/SCID female mice anticipating that these invasive cells would most likely form tumors. None of the mice developed tumors by week 20. From such results it was clear that two mutations in a cell were not enough to form tumors in this context. In this *in-vivo* system cell invasiveness was not enough for tumor formation (or progression to metastasis) and that *in-vitro* invasion assays cannot be surrogates for the tumorigenic or metastasis process.

The post-invasion cell population of GeCKO v2 library transduced R273C cells was injected into the mice to see if a strongly invasive p53 mutant like R273C could drive rapid cell transformation and tumor formation. No tumors were observed in this case. In order to develop cells carrying multiple mutations (at least three mutations including p53 mutation), we transduced the MCF 10A p53 mutant R273C cells with the GeCKO v2 library at high MOI (HMOI>4) such that majority of the cell fraction received multiple sgRNAs and subsequently multiple mutations in addition to p53 R273C. Both pre and post-invasion R273C libA and libB HMOI cell populations were injected orthotopically into the mammary fat pads of the female NOD/SCID mice. In such a case also, tumors did not develop by week 16.

One must keep in mind that we have generated the cell lines with multiple mutations using a CRISPR based knock-out library. Hence, the mutated genes would all be possible tumor suppressors and not oncogenes. Cancer is a multistep process and requires the interplay between loss-of-function of tumor suppressors and gain-of-function of oncogenes. Amplification of known oncogenes like EGFR or MYC is observed in the aggressive TNBC tumors. The metastatic TNBC cell-lines such as MDA-MB-468 and MDA-MB-436 that harbor mutations in *TP53* also contain amplification of *EGFR* and *MYC* respectively. Previous work by Robert Weinberg's group have shown that oncogenes are essential to cause transformation of cells and that the double mutant mice with loss of p53 function and transgenic for MYC show

early onset of tumors. We thus, generated cell-lines with combination of mutations in tumor suppressors and overexpression of EGFR or MYC that would possibly turn cells tumorigenic. Large primary tumors were observed in the NOD/SCID mice injected with R273C MYC libB and R273C EGFR libB by week 9 but the tumor cells did not metastasize. Single-cell sequencing from these tumors will illuminate the gene mutation combinations that led to tumor formation.

Though cells with a combination of mutations in tumor suppressors and the overexpression of an oncogene turned tumorigenic, accumulation of more mutations are required for these tumorigenic cells to metastasize. The human cells need accumulation of more genetic events for neoplasm formation than the murine cells. The tumorigenic conversion of normal mouse fibroblasts required disruption of two signaling pathways but for human fibroblasts six pathways needed to be disrupted (Rangarajan 2004). The requirement for tumorigenic transformation of human cells is cell-type dependent. While the human mammary cells needed activation of PI3K, Raf and Ral-GEF, the immortalized human fibroblasts required the activation of only Raf and Ral-GEF. This work reflected the complex multistep tumorigenesis process in two different species, mice and humans (Rangarajan 2004).

Though we are using a mouse xenograft model and injecting the experimental cells in high concentration with Matrigel into the orthotopic mammary fat pads of the female NOD/SCID mice, the total number of cells bearing the genetic lesions is very small. The timeline for observation of the xenograft bearing mice could also affect the results and it is possible that our experimental mice have very long periods of tumor latency.

For studying the process of metastasis, tumorigenicity of the cells is a pre-requisite. Once the cells undergo tumorigenic transformation followed by metastasis, this *in-vivo* system could provide valuable information on the specific gene combinations driving the

complex process of tumor metastasis. The extraction of primary tumors and metastatic tumors from specific organs and their sequencing can elucidate the process of tumor evolution in cancer, the genes involved in tissue tropism during metastasis and address the controversial topic of whether metastasis is an early or a late event in breast cancer.

References

Al-Kuraya, Khawla, Peter Schraml, Joachim Torhorst, Coya Tapia, Boriana Zaharieva, Hedvika Novotny, Hanspeter Spichtin et al. "Prognostic relevance of gene amplifications and coamplifications in breast cancer." *Cancer research* 64, no. 23 (2004): 8534-8540.

Attardi, Laura D., and Lawrence A. Donehower. "Probing p53 biological functions through the use of genetically engineered mouse models." *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 576, no. 1-2 (2005): 4-21.

Chavez, Kathryn J., Sireesha V. Garimella, and Stanley Lipkowitz. "Triple negative breast cancer cell lines: one tool in the search for better treatment of triple negative breast cancer." *Breast disease* 32, no. 1-2 (2010): 35.

Elenbaas, Brian, Lisa Spirio, Frederick Koerner, Mark D. Fleming, Drazen B. Zimonjic, Joana Liu Donaher, Nicholas C. Popescu, William C. Hahn, and Robert A. Weinberg. "Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells." *Genes & development* 15, no. 1 (2001): 50-65.

Hahn, William C., Christopher M. Counter, Ante S. Lundberg, Roderick L. Beijersbergen, Mary W. Brooks, and Robert A. Weinberg. "Creation of human tumour cells with defined genetic elements." *Nature* 400, no. 6743 (1999): 464.

Ito, Ryoji, Takeshi Takahashi, Ikumi Katano, and Mamoru Ito. "Current advances in humanized mouse models." *Cellular & molecular immunology* 9, no. 3 (2012): 208.

Kim, Gwangil, Maria Ouzounova, Ahmed A. Quraishi, April Davis, Nader Tawakkol, Shawn G. Clouthier, Fayaz Malik et al. "SOCS3-mediated regulation of inflammatory cytokines in PTEN and p53 inactivated triple negative breast cancer model." *Oncogene* 34, no. 6 (2015): 671.

Land, Hartmut, Luis F. Parada, and Robert A. Weinberg. "Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes." *Nature* 304, no. 5927 (1983): 596.

Liao, D. J., and R. B. Dickson. "c-Myc in breast cancer." *Endocrine-related cancer* 7, no. 3 (2000): 143-164.

Parada, Luis F., Hartmut Land, Robert A. Weinberg, David Wolf, and Varda Rotter. "Cooperation between gene encoding p53 tumour antigen and ras in cellular transformation." *Nature*312, no. 5995 (1984): 649.

Rangarajan, Annapoorni, Sue J. Hong, Annie Gifford, and Robert A. Weinberg. "Species-and cell type-specific requirements for cellular transformation." *Cancer cell* 6, no. 2 (2004): 171-183.

CHAPTER 6

SUMMARY AND FUTURE DIRECTIONS

This dissertation sheds light on the dominant negative 'neo-morphic' activity of the ten missense mutations in *TP53* prevalent in breast tumors and other collaborating somatic co-driver mutations in tumor suppressor genes that drive cell invasion and aggressiveness in the context of breast epithelial cells.

Summary of the results

Many missense mutations occur in the *TP53* tumor suppressor gene concentrating in the DNA binding domain (DBD) of the protein. We hypothesized that the different missense mutations in the DBD affect the sequence specific binding of the mutant p53 transcription factor and hence, the transcriptional activity and the downstream pathways resulting in different phenotypes. In chapter 2, we described the detailed molecular profiling of the ten different p53 missense mutant proteins over-expressed in the non-transformed MCF 10A mammary epithelial cells that revealed 'neo-morphic' activities of the mutant proteins based on the hallmarks of cancer. These important hallmarks included ability to survive and proliferate under growth factor starvation, resistance to cell apoptosis and anoikis, cell migration and invasion through Matrigel matrix and cell polarity change in mammospheres formed in 3D matrigel matrix (Chapter 2, Fig 2.8). The p53 mutants R248W, R273C and Y220C clustered as the most aggressive mutants whereas mutants G245S, R273H and Y234C were found to be least aggressive. All the mutants demonstrated a spectrum of phenotypes.

The integrated analysis of ChIP seq showed distinct promoter binding profiles of the different p53 mutants as transcription factors and RNA seq highlighted the activated pathways such as the focal adhesion pathway as one of the dysregulated pathways in the invasive mutants (Chapter 2,

Fig 2.9). We saw a good correlation between the observed phenotypes and the enriched pathways. Enrichment and model-based pathway analysis revealed the key biological pathways associated with metabolism, axon guidance and focal adhesion pathways during invasion of cells (Chapter 2, Fig 2.9). These results implied that non-canonical transcriptional activity may contribute to the phenotypic heterogeneity seen in the *TP53* mutated breast tumors. The thorough phenotypic comparison of the MCF 10A cells expressing different missense mutant p53 proteins provide information on how the *TP53* mutations differentially impact cancer phenotypes.

The mutations in *TP53* are early events in breast cancer progression. They may contribute to the breast tumor initiation but are not enough to drive breast cancer progression. Several other somatic mutations co-appear with mutation in *TP53* that are less prevalent but biologically significant and contribute to progression of breast cancer. We introduced the term ‘co-drivers’ of mutant p53 proteins that work in synergy or in addition to mutant p53 to drive the process of tumorigenesis. In chapter 3, we demonstrated that the deletion of another tumor suppressor gene *PTEN* acts as the co-driver of the missense mutant *TP53* to drive breast epithelial cell invasion (Chapter 3, Fig 3.5). Homozygous deletion of *PTEN* is known to co-occur with *TP53* mutations in basal-like breast cancer. Shut-down of the PTEN protein using shRNA and knocking out the *PTEN* gene using the CRISPR-Cas9 system in the weakly invasive MCF 10A p53 mutant cells (G245S, R273H and Y234C) resulted in increased cell invasion. The combination of *PTEN* deletion and missense mutant p53 also transformed the cells through anchorage independent growth as seen in the soft agar assays (Chapter 3, Fig 3.11). This study provided the proof-of-principle for the somatically mutated ‘co-driver’ genes of missense mutant *TP53* that work in synergy to drive increased cell invasion or aggressiveness in mammary epithelial cells. The MCF 10A cells with mutant p53 and deletion of PTEN as the co-driver served as a positive control for the functional genomic pipeline that we set up for the screening of the possible co-drivers of missense mutant p53 on a genome wide scale.

Powerful tools like Next Generation Sequencing (NGS) has uncovered the mutational landscape of cancer tumors. These somatic mutations include a few recurrently mutated drivers, many less-frequently mutated co-drivers and most passenger mutations. Separation of these biologically relevant co-drivers from the neutral passenger mutations is a daunting task. Chapter 4 discussed a CRISPR based functional genomic approach used effectively for screening a *phenotype* and discovering the *genotype* that caused it. As most of the somatic mutations co-existing with *TP53* mutation are likely to be non-functional passenger mutations, this functional screen sieved out the possible functionally relevant 'co-drivers' of mutant *TP53*. We screened the weakly invasive MCF 10A p53 Y234C and the highly invasive p53 R273C cells with the genome wide GeCKO v2 lentivirus library for increased cell invasion and found some known 'cancer related genes' such as *BCL2, PTEN, EPHB2* and a number of lesser known genes with plausible role in promoting breast cancer aggressiveness as our top hits from the screening. The two different *TP53* contexts enriched different set of sgRNAs in the post-invasion cell population. From each screen, different *'co-driver'* genes appeared in the p53 Y234C and p53 R273C cells to drive cell invasiveness (Chapter 4, Fig 4.8). The top hits chosen from the p53 Y234C cell screening were *GCNT1, FGD6* and *ADAM15* while *PACS1* and *SDF4* were the most enriched hits in the p53 R273C screen. These mutation combinations are likely to affect breast cancer tumor heterogeneity.

We were curious to know if the post-invasion cell population of Y234C and R273C cells with combination of co-drivers could evolve in an *in-vivo* system and generate tumors. We anticipated that the cell population could drive tumorigenesis in mice as the cells injected were already invasive. From chapter 2 to chapter 4, we had all the experimental results from *in-vitro* (cell-based) studies. In chapter 5 we presented all the results from the *in-vivo* study, where we established a mouse xenograft model to observe tumor formation and metastasis (if primary tumors formed). We observed that the post-invasion cell population of 10A Y234C and R273C cells with one co-driver (transduction with low MOI) or with more than two additional co-drivers (transduction with high MOI) could not form tumors in the immunodeficient NOD/SCID mice (Chapter 5, Fig 5.3). The results implied that more than three or four mutations are required to

make cells tumorigenic in this context and a combination of mutations in both tumor suppressor genes and oncogenes may promote tumor formation instead of mutations only in tumor suppressors (GeCKO v2 is a knock-out library). For the tumors to metastasize, accumulation of more mutations is required as seen in the metastatic breast cancer cell-lines for example, MDA-MB-231 (positive control) that harbors 5-6 driver mutations. Breast tumors from patients harboring *TP53* mutation also contain hundreds of other somatic mutations. We over-expressed the known oncogenes, EGFR or MYC, commonly observed in TNBC cell-lines, in library transduced MCF 10A p53 mutant R273C (at high MOI) cell-lines and injected into the mammary fat pads of the NOD/SCID mice. As an appropriate positive control, we also generated R273C cells with *PTEN* deletion and over-expression of either EGFR or MYC. We expect to see tumors with the combination of mutant tumor suppressor and overexpressed oncogenes in addition to different co-drivers. Large primary tumors formed in the mice injected with R273C library transduced (at high MOI) and overexpressing either EGFR or MYC MCF 10A cells by the 9^{th} week. The mice did not show metastasis. We would be performing single cell sequencing to reveal the mutation combinations resulting in tumorigenesis.

In summary, we performed an extensive molecular phenotyping of the MCF 10A cells expressing the ten most frequent missense *TP53* mutations occurring in breast cancer based on five hallmarks of cancer that revealed heterogeneous neo-morphic phenotypes of missense mutant p53 proteins. The ChIP and RNA seq studies provided a deeper knowledge of the transcriptomics of the mutant proteins and the resultant pathways shaping the phenotype. To our knowledge, this is the first study to present an extensive comparative study on phenotypes imparted by a total of ten p53 missense mutations and a spectrum of phenotypes generated by these mutants. The functional genomics approach to differentiate between the non-functional passenger mutations and the relevant functional drivers proved to be an efficient method to identify the co-drivers of mutant p53 in different cellular contexts. Overall, we found that p53 mutations alone are insufficient to promote transformation of normal-like MCF 10A mammary epithelial cells. Other genes (tumor suppressors or oncogenes) bearing somatic mutations and designated as the ‘co-

drivers' collaborate with the driver mutant *TP53* to accelerate breast cancer progression. And even a larger number of mutations are required for the cells to achieve full malignancy as deduced from the *in-vivo* studies.

Impact of the study

TP53 mutations are most frequent (88%) in the TNBC which is the most aggressive breast cancer subtype. Our results from the molecular profiling of the missense mutant p53 proteins demonstrate their neo-morphic activities and shed light on the phenotypic heterogeneity observed in TNBC tumors harboring different *TP53* mutations.

The elucidation of different unique co-drivers for different missense mutant p53 can be informative for developing potential personalized drug target genes and pathways for TNBC patients with specific *TP53* mutation.

A function-based pipeline as opposed to predictive computational methods makes an efficient pipeline to screen for the biologically relevant co-drivers. This method can be applied to screen for other cell phenotypes (proliferation, apoptosis resistance, drug resistance) and different cancer context.

Limitations of the study

The phenotypic effects of missense mutant p53 proteins have been studied in only one normal-like non-transformed mammary epithelial cell line MCF 10A. Another breast epithelial cell line, the non-transformed HMECs immortalized with expression of *hTERT* gene gave very inconsistent

phenotypic results with the expression of missense mutant p53 proteins. Hence, we chose the MCF 10A cells for our study which provided consistent results throughout the experiments.

In the MCF 10A cells the mutant p53 proteins have been over-expressed using a lentiviral expression vector and there is no manipulation at the genome level which does not truly model the conditions found in the breast tumors.

For the co-driver screening purpose, we have only used a CRISPR based gene knock-out library (GeCKO v2) and thus, considered only other tumor suppressor genes as the 'co-drivers' of the missense mutant *TP53*.

Future directions

Validation of the co-drivers as drug targets

Direct targeting of the co-drivers would not be possible here as we used a knock-out library for the screening. The dysregulated pathways for example, genes participating in the focal adhesion pathway can be targeted by available drugs or gene editing to abrogate cell invasion.

Mice models with specific combination of *TP53* mutations and unique combination of co-drivers forming primary tumors and eventual metastasis can serve as pre-clinical models for developing alternative targeted therapies in the context of otherwise difficult to target mutant p53.

Study of how different co-drivers promote invasion

RNA sequencing of the post-invasion MCF 10A p53 Y234C and p53 R273C cells with co-driver combinations can determine the altered transcriptome of the cells and molecular pathways that promote cell invasion in the different *TP53* mutation background (for example, p53 Y234C GCNT1$^{-/-}$ or p53 R273C PACS1$^{-/-}$ cells).

Screening of oncogenes as co-drivers of missense mutant p53

The p53 mutant MCF 10A cells can be screened with CRISPR based activation or repression libraries (commercially available) or with a custom-made library containing small gRNAs (sgRNAs) for activation and repression of genes in a given library. This would help in determining the combinations of oncogenes or oncogenes and mutated tumor suppressors genes as co-drivers that collaborate with mutant p53 to drive cell invasion or any other hallmarks of cancer.

Transduction of the p53 mutant cells at high multiplicity of infection (HMOI)

The screening for co-driver sgRNA combinations by using conventional approaches require multiple rounds of successive clonal expansion and screens (*i.e*., “stepwise clonal screen”) that can be time consuming and cannot accommodate the complexity of all possible gene combinations as found in tumors (Fig 6.1). The stepwise screen may also suffer from a low discovery rate and a new selection marker is needed at each screening step. Moreover, mutations that work singly may differ from mutations that work in combinations.

Multiple mutations can be introduced in the p53 mutant MCF 10A cells by transducing the cells at high MOI instead of maintaining the transduction conditions at low MOI as

done for the screening of mutant p53 co-drivers (chapter 4). In this way the cells can acquire two or more mutations (sgRNAs) at a time in addition to the mutant p53. Such mutation combinations contributing to increased cell invasion can be decoded by single cell sequencing. A “pooled shotgun screen” with transduction maintained at high MOI (MOI=4 or 5), the assessment of co-existing perturbations in each invasive or metastatic cell can be made. Since there is no technology available for supporting such screens, we are developing a novel high-throughput amplicon droplet sequencing platform in collaboration with Dr. Nikkhah, a microfluidics engineer at ASU. Unlike conventional droplet sequencing method developed for reading whole genome or transcriptome information from a single cell, this amplicon-targeted platform can PCR-amplify and quantify multiple perturbagens in a single cell, specifically for genome wide co-driver screens.

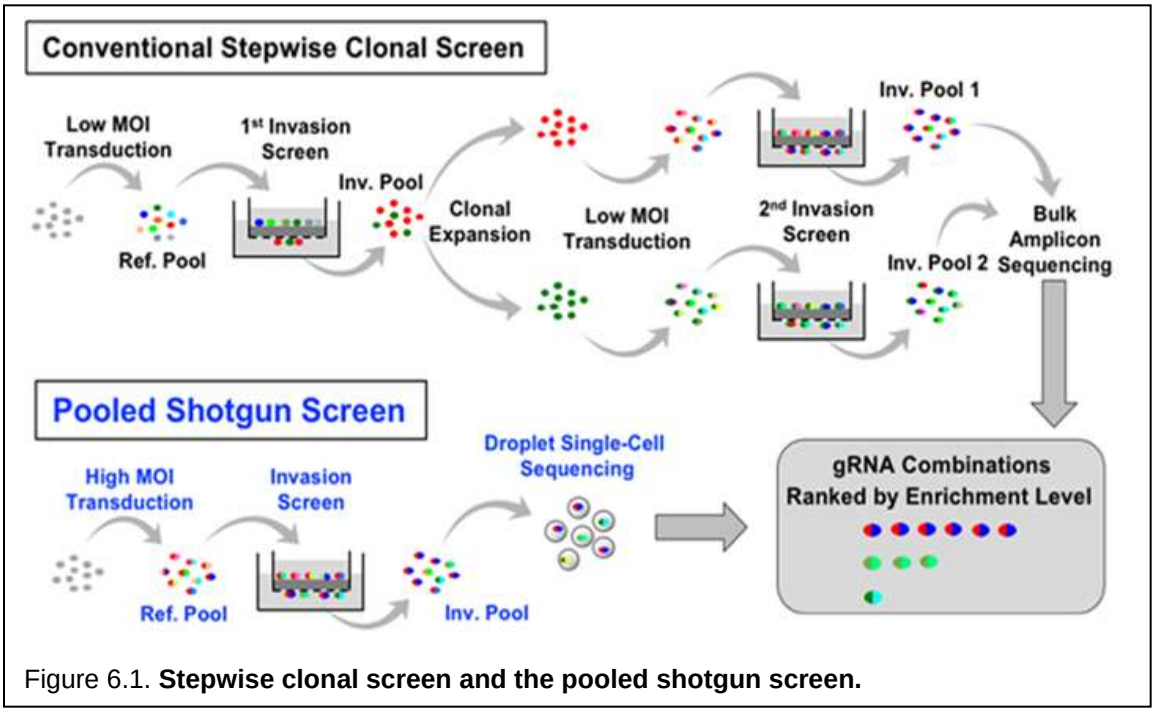


Figure 6.1. **Stepwise clonal screen and the pooled shotgun screen.**

Genetically altered mouse model for tumor formation and metastasis

Determination of co-driver mutation combinations from the single cell sequencing results and generating genetically manipulated mice with such mutation combinations can determine the tumor spectrum and their consequent metastatic capabilities. Such a model could help to study genomic evolution of the tumors over time and act as a pre-clinical model to discover the druggable targets or pathways to contain the tumor growth.

REFERENCES

Aas, Turid, Anne-Lise Børresen, Stephanie Geisler, Birgitte Smith-Sørensen, Hilde Johnsen, Jan E. Varhaug, Lars A. Akslen, and Per E. Lønning. "Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients." *Nature medicine* 2, no. 7 (1996): 811.

Ademuyiwa, Foluso O., Yu Tao, Jingqin Luo, Katherine Weilbaecher, and Cynthia X. Ma. "Differences in the mutational landscape of triple-negative breast cancer in African Americans and Caucasians." *Breast cancer research and treatment* 161, no. 3 (2017): 491-499.

Alexandrova, Evguenia M., Safia A. Mirza, Sulan Xu, Ramona Schulz-Heddergott, Natalia D. Marchenko, and Ute M. Moll. "p53 loss-of-heterozygosity is a necessary prerequisite for mutant p53 stabilization and gain-of-function in vivo." *Cell death & disease* 8, no. 3 (2017): e2661.

Al-Kuraya, Khawla, Peter Schraml, Joachim Torhorst, Coya Tapia, Boriana Zaharieva, Hedvika Novotny, Hanspeter Spichtin et al. "Prognostic relevance of gene amplifications and coamplifications in breast cancer." *Cancer research* 64, no. 23 (2004): 8534-8540.

Ashcroft, Margaret, and Karen H. Vousden. "Regulation of p53 stability." *Oncogene* 18, no. 53 (1999): 7637.

Attardi, Laura D., and Lawrence A. Donehower. "Probing p53 biological functions through the use of genetically engineered mouse models." *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 576, no. 1-2 (2005): 4-21.

Aubrey, Brandon J., Ana Janic, Yunshun Chen, Catherine Chang, Elizabeth C. Lieschke, Sarah T. Diepstraten, Andrew J. Kueh et al. "Mutant TRP53 exerts a target gene-selective dominant-negative effect to drive tumor development." *Genes & development* 32, no. 21-22 (2018): 1420-1429.

Bailey, Matthew H., Collin Tokheim, Eduard Porta-Pardo, Sohini Sengupta, Denis Bertrand, Amila Weerasinghe, Antonio Colaprico et al. "Comprehensive characterization of cancer driver genes and mutations." *Cell* 173, no. 2 (2018): 371-385.

Bailey, Timothy L., Mikael Boden, Fabian A. Buske, Martin Frith, Charles E. Grant, Luca Clementi, Jingyuan Ren, Wilfred W. Li, and William S. Noble. "MEME SUITE: tools for motif discovery and searching." *Nucleic acids research* 37, no. suppl_2 (2009): W202-W208.

Baker, Suzanne J., Eric R. Fearon, Janice M. Nigro, A. C. Preisinger, J. M. Jessup, D. H. Ledbetter, D. F. Barker, Y. Nakamura, R. White, and B. Vogelstein. "Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas." *Science* 244, no. 4901 (1989): 217-221.

Banerji, Shantanu, Kristian Cibulskis, Claudia Rangel-Escareno, Kristin K. Brown, Scott L. Carter, Abbie M. Frederick, Michael S. Lawrence et al. "Sequence analysis of mutations and translocations across breast cancer subtypes." *Nature* 486, no. 7403 (2012): 405.

Beckerman, Rachel, and Carol Prives. "Transcriptional regulation by p53." *Cold Spring Harbor perspectives in biology*2, no. 8 (2010): a000935.

Beckmann, M. W., Dieter Niederacher, Hans-Georg Schnürch, Barry A. Gusterson, and Hans Georg Bender. "Multistep carcinogenesis of breast cancer and tumour heterogeneity." *Journal of molecular medicine* 75, no. 6 (1997): 429-439.

Bergh, Jonas, Torbjörn Norberg, Sigrid Sjögren, Anders Lindgren, and Lars Holmberg. "Complete sequencing of the p53 gene provides prognostic information in breast cancer patients, particularly in relation to adjuvant systemic therapy and radiotherapy." *Nature medicine* 1, no. 10 (1995): 1029.

Bertheau, Philippe, Jacqueline Lehmann-Che, Mariana Varna, Anne Dumay, Brigitte Poirot, Raphaël Porcher, Elisabeth Turpin et al. "p53 in breast cancer subtypes and new insights into response to chemotherapy." *The Breast* 22 (2013): S27-S29.

Bianchini, Giampaolo, Justin M. Balko, Ingrid A. Mayer, Melinda E. Sanders, and Luca Gianni. "Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease." *Nature reviews Clinical oncology* 13, no. 11 (2016): 674.

Bieging, Kathryn T., Stephano Spano Mello, and Laura D. Attardi. "Unravelling mechanisms of p53-mediated tumour suppression." *Nature Reviews Cancer* 14, no. 5 (2014): 359.

Billant, Olivier, Alice Léon, Solenn Le Guellec, Gaëlle Friocourt, Marc Blondel, and Cécile Voisset. "The dominant-negative interplay between p53, p63 and p73: A family affair." *Oncotarget* 7, no. 43 (2016): 69549.

Birch, Jillian M., Valerie Blair, Anna M. Kelsey, D. Gareth Evans, Martin Harris, Karen J. Tricker, and Jennifer M. Varley. "Cancer phenotype correlates with constitutional TP53 genotype in families with the Li–Fraumeni syndrome." *Oncogene* 17, no. 9 (1998): 1061.

Blandino, Giovanni, Arnold J. Levine, and Moshe Oren. "Mutant p53 gain of function: differential effects of different p53 mutants on resistance of cultured cells to chemotherapy." *Oncogene* 18, no. 2 (1999): 477.

Bode, Ann M., and Zigang Dong. "Post-translational modification of p53 in tumorigenesis." *Nature Reviews Cancer*4, no. 10 (2004): 793.

Boettcher, Michael, and Michael T. McManus. "Choosing the right tool for the job: RNAi, TALEN, or CRISPR." *Molecular cell*58, no. 4 (2015): 575-585.
Børresen-Dale, Anne-Lise. "TP53 and breast cancer." *Human mutation* 21, no. 3 (2003): 292-300.

Brasacchio, Daniella, Amber E. Alsop, Tahereh Noori, Mariam Lufti, Sweta Iyer, Kaylene J. Simpson, Phillip I. Bird, Ruth M. Kluck, Ricky W. Johnstone, and Joseph A. Trapani. "Epigenetic control of mitochondrial cell death through PACS1-mediated regulation of BAX/BAK oligomerization." *Cell death and differentiation* 24, no. 6 (2017): 961.

Brosh, Ran, and Varda Rotter. "When mutants gain new powers: news from the mutant p53 field." *Nature reviews cancer* 9, no. 10 (2009): 701.

Burdelski, Christoph, Michael Fitzner, Claudia Hube-Magg, Martina Kluth, Asmus Heumann, Ronald Simon, Till Krech et al. "Overexpression of the A Disintegrin and Metalloproteinase ADAM15 is linked to a Small but Highly Aggressive Subset of Prostate Cancers." *Neoplasia* 19, no. 4 (2017): 279-287.

Burrell, Rebecca A., Nicholas McGranahan, Jiri Bartek, and Charles Swanton. "The causes and consequences of genetic heterogeneity in cancer evolution." *Nature* 501, no. 7467 (2013): 338.

Cancer Genome Atlas Network. "Comprehensive molecular portraits of human breast tumours." *Nature* 490, no. 7418 (2012): 61.

Canver, Matthew C., Daniel E. Bauer, Abhishek Dass, Yvette Y. Yien, Jacky Chung, Takeshi Masuda, Takahiro Maeda, Barry H. Paw, and Stuart H. Orkin. "Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells." *Journal of Biological Chemistry*289, no. 31 (2014): 21312-21324.

Cao, Zheng, Theodore Livas, and Natasha Kyprianou. "Anoikis and EMT: Lethal" liaisons" during cancer progression." *Critical Reviews™ in Oncogenesis* 21, no. 3-4 (2016).

Carey, Lisa, Eric Winer, Giuseppe Viale, David Cameron, and Luca Gianni. "Triple-negative breast cancer: disease entity or title of convenience?." *Nature reviews Clinical oncology* 7, no. 12 (2010): 683.

Castro-Giner, Francesc, Peter Ratcliffe, and Ian Tomlinson. "The mini-driver model of polygenic cancer evolution." *Nature Reviews Cancer* 15, no. 11 (2015): 680.

Chalhoub, Nader, and Suzanne J. Baker. "PTEN and the PI3-kinase pathway in cancer." *Annual Review of Pathological Mechanical Disease* 4 (2009): 127-150.

Chavez, Kathryn J., Sireesha V. Garimella, and Stanley Lipkowitz. "Triple negative breast cancer cell lines: one tool in the search for better treatment of triple negative breast cancer." *Breast disease* 32, no. 1-2 (2010): 35.

Chen, Sidi, Neville E. Sanjana, Kaijie Zheng, Ophir Shalem, Kyungheon Lee, Xi Shi, David A. Scott et al. "Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis." *Cell* 160, no. 6 (2015): 1246-1260.

Cho, Yunje, Svetlana Gorina, Philip D. Jeffrey, and Nikola P. Pavletich. "Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations." *Science* 265, no. 5170 (1994): 346-355.

Chow, Ryan D., and Sidi Chen. "Cancer CRISPR screens in vivo." *Trends in cancer* 4, no. 5 (2018): 349-358.

Crevenna, Alvaro H., Birgit Blank, Andreas Maiser, Derya Emin, Jens Prescher, Gisela Beck, Christine Kienzle et al. "Secretory cargo sorting by Ca2+-dependent Cab45 oligomerization at the trans-Golgi network." *J Cell Biol* 213, no. 3 (2016): 305-314.

Davidoff, Andrew M., Billie-Jo M. Kerns, J. Dirk Iglehart, and Jeffrey R. Marks. "Maintenance of p53 alterations throughout breast cancer progression." *Cancer Research* 51, no. 10 (1991): 2605-2610.

De Vries, Annemieke, Elsa R. Flores, Barbara Miranda, Harn-Mei Hsieh, Conny Th M. van Oostrom, Julien Sage, and Tyler Jacks. "Targeted point mutations of p53 lead to dominant-negative inhibition of wild-type p53 function." *Proceedings of the National Academy of Sciences* 99, no. 5 (2002): 2948-2953.

Debnath, Jayanta, Senthil K. Muthuswamy, and Joan S. Brugge. "Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures." *Methods* 30, no. 3 (2003): 256-268.

Derksen, Patrick WB, Xiaoling Liu, Francis Saridin, Hanneke van der Gulden, John Zevenhoven, Bastiaan Evers, Judy R. van Beijnum et al. "Somatic inactivation of E-cadherin and p53 in mice leads to metastatic lobular mammary carcinoma through induction of anoikis resistance and angiogenesis." *Cancer cell* 10, no. 5 (2006): 437-449.

Di Agostino, Silvia, Sabrina Strano, Velia Emiliozzi, Valentina Zerbini, Marcella Mottolese, Ada Sacchi, Giovanni Blandino, and Giulia Piaggio. "Gain of function of mutant p53: the mutant p53/NF-Y protein complex reveals an aberrant transcriptional mechanism of cell cycle regulation." *Cancer cell* 10, no. 3 (2006): 191-202.

Dittmer, Dirk, Sibani Pati, Gerard Zambetti, Shelley Chu, Angelika K. Teresky, Mary Moore, Cathy Finlay, and Arnold J. Levine. "Gain of function mutations in p53." *Nature genetics* 4, no. 1 (1993): 42.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. "STAR: ultrafast universal RNA-seq aligner." *Bioinformatics* 29, no. 1 (2013): 15-21.

Donehower, Lawrence A., Michele Harvey, Betty L. Slagle, Mark J. McArthur, Charles A. Montgomery Jr, Janet S. Butel, and Allan Bradley. "Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours." *Nature* 356, no. 6366 (1992): 215.

Elenbaas, Brian, Lisa Spirio, Frederick Koerner, Mark D. Fleming, Drazen B. Zimonjic, Joana Liu Donaher, Nicholas C. Popescu, William C. Hahn, and Robert A. Weinberg. "Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells." *Genes & development* 15, no. 1 (2001): 50-65.

Eliyahu, Daniel, Avraham Raz, Peter Gruss, David Givol, and Moshe Oren. "Participation of p53 cellular tumour antigen in transformation of normal embryonic cells." *Nature* 312, no. 5995 (1984): 646.

Finlay, C. A., P. W. Hinds, T. H. Tan, D. Eliyahu, M. Oren, and A. J. Levine. "Activating mutations for transformation by p53 produce a gene product that forms an hsc70-p53 complex with an altered half-life." *Molecular and cellular biology* 8, no. 2 (1988): 531-539.

Finlay, Cathy A., Philip W. Hinds, and Arnold J. Levine. "The p53 proto-oncogene can act as a suppressor of transformation." *Cell* 57, no. 7 (1989): 1083-1093.

Fischer, M. "Census and evaluation of p53 target genes." *Oncogene* 36, no. 28 (2017): 3943.

Fisher, R., L. Pusztai, and C. Swanton. "Cancer heterogeneity: implications for targeted therapeutics." *British journal of cancer*108, no. 3 (2013): 479.

Freed-Pastor, William A., and Carol Prives. "Mutant p53: one name, many proteins." *Genes & development* 26, no. 12 (2012): 1268-1286.

Fujii, Hiroaki, Carla Marsh, Paul Cairns, David Sidransky, and Edward Gabrielson. "Genetic divergence in the clonal evolution of breast cancer." *Cancer Research* 56, no. 7 (1996): 1493-1497.

Gao, Ruli, Alexander Davis, Thomas O. McDonald, Emi Sei, Xiuqing Shi, Yong Wang, Pei-Ching Tsai et al. "Punctuated copy number evolution and clonal stasis in triple-negative breast cancer." *Nature genetics* 48, no. 10 (2016): 1119.

Georgescu, Maria-Magdalena, Kathrin H. Kirsch, Tsuyoshi Akagi, Tomoyuki Shishido, and Hidesaburo Hanafusa. "The tumor-suppressor activity of PTEN is regulated by its carboxyl-terminal region." *Proceedings of the National Academy of Sciences* 96, no. 18 (1999): 10182-10187.

Gilbert, Luke A., Max A. Horlbeck, Britt Adamson, Jacqueline E. Villalta, Yuwen Chen, Evan H. Whitehead, Carla Guimaraes et al. "Genome-scale CRISPR-mediated control of gene repression and activation." *Cell* 159, no. 3 (2014): 647-661.

Gilmore, A. P. "Anoikis." (2005): 1473.

Goh, Amanda M., Cynthia R. Coffill, and David P. Lane. "The role of mutant p53 in human cancer." *The Journal of pathology*223, no. 2 (2011): 116-126.

Gottlieb, Eyal, and Karen H. Vousden. "p53 regulation of metabolic pathways." *Cold Spring Harbor perspectives in biology* 2, no. 4 (2010): a001040.

Gray, Joe, and Brian Druker. "Genomics: the breast cancer landscape." *Nature* 486, no. 7403 (2012): 328.

Greaves, Mel, and Carlo C. Maley. "Clonal evolution in cancer." *Nature* 481, no. 7381 (2012): 306.

Greenblatt, M. S., William P. Bennett, M. Hollstein, and C. C. Harris. "Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis." *Cancer research* 54, no. 18 (1994): 4855-4878.

Greenman, Christopher, Philip Stephens, Raffaella Smith, Gillian L. Dalgliesh, Christopher Hunter, Graham Bignell, Helen Davies et al. "Patterns of somatic mutation in human cancer genomes." *Nature* 446, no. 7132 (2007): 153.

Guha, Tanya, and David Malkin. "Inherited TP53 mutations and the Li–Fraumeni syndrome." *Cold Spring Harbor perspectives in medicine* 7, no. 4 (2017): a026187.

Hafner, Antonina, Martha L. Bulyk, Ashwini Jambhekar, and Galit Lahav. "The multiple mechanisms that regulate p53 activity and cell fate." *Nature reviews. Molecular cell biology*(2019).

Hagisawa, Shigeru, Chikara Ohyama, Toshiko Takahashi, Mareyuki Endoh, Takuya Moriya, Jun Nakayama, Yoichi Arai, and Minoru Fukuda. "Expression of core 2 β1, 6-N-acetylglucosaminyltransferase facilitates prostate cancer progression." *Glycobiology* 15, no. 10 (2005): 1016-1024.

Hahn, William C., Christopher M. Counter, Ante S. Lundberg, Roderick L. Beijersbergen, Mary W. Brooks, and Robert A. Weinberg. "Creation of human tumour cells with defined genetic elements." *Nature* 400, no. 6743 (1999): 464.

Hainaut, Pierre, and Gerd P. Pfeifer. "Somatic TP53 mutations in the era of genome sequencing." *Cold Spring Harbor Perspectives in Medicine* 6, no. 11 (2016): a026179.
Hanahan, Douglas, and Robert A. Weinberg. "Hallmarks of cancer: the next generation." *cell* 144, no. 5 (2011): 646-674.

Hanel, W., N. Marchenko, S. Xu, S. Xiaofeng Yu, W. Weng, and U. Moll. "Two hot spot mutant p53 mouse models display differential gain of function in tumorigenesis." *Cell death and differentiation* 20, no. 7 (2013): 898.

Harvey, Michele, Hannes Vogel, Danna Morris, Allan Bradley, Alan Bernstein, and Lawrence A. Donehower. "A mutant p53 transgene accelerates tumour development in heterozygous but not nullizygous p53–deficient mice." *Nature genetics* 9, no. 3 (1995): 305.

Heinz, Sven, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh, and Christopher K. Glass. "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." *Molecular cell* 38, no. 4 (2010): 576-589.

Hiles, Guadalupe Lorenzatti, Amanda Bucheit, John R. Rubin, Alexandra Hayward, Angelica L. Cates, Kathleen C. Day, Layla El-Sawy et al. "ADAM15 is functionally associated with the metastatic progression of human bladder cancer." *PloS one* 11, no. 3 (2016): e0150138.

Hollestelle, Antoinette, Jord HA Nagel, Marcel Smid, Suzanne Lam, Fons Elstrodt, Marijke Wasielewski, Ser Sue Ng et al. "Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines." *Breast cancer research and treatment* 121, no. 1 (2010): 53-64.

Hollstein, Monica, David Sidransky, Bert Vogelstein, and Curtıs C. Harris. "p53 mutations in human cancers." *Science*253, no. 5015 (1991): 49-53.

Huang, Wenhuan, Cheng Chen, Zhongheng Liang, Junlu Qiu, Xinxin Li, Xiang Hu, Shuanglin Xiang, Xiaofeng Ding, and Jian Zhang. "AP-2α inhibits hepatocellular carcinoma cell growth and migration." *International journal of oncology* 48, no. 3 (2016): 1125-1134.

Ikeda, Osamu, Yuichi Sekine, Akihiro Mizushima, Misa Nakasuji, Yuto Miyasaka, Chikako Yamamoto, Ryuta Muromoto et al. "Interactions of STAP-2 with Brk and STAT3 participate in cell growth of human breast cancer cells." *Journal of Biological Chemistry* 285, no. 49 (2010): 38093-38103.

Ilić, Duško, Eduardo AC Almeida, David D. Schlaepfer, Paul Dazin, Shinichi Aizawa, and Caroline H. Damsky. "Extracellular matrix survival signals transduced by focal adhesion kinase suppress p53-mediated apoptosis." *The Journal of cell biology* 143, no. 2 (1998): 547-560.

Iranzo, Jaime, Iñigo Martincorena, and Eugene V. Koonin. "Cancer-mutation network and the number and specificity of driver mutations." *Proceedings of the National Academy of Sciences* 115, no. 26 (2018): E6010-E6019.

Ito, Ryoji, Takeshi Takahashi, Ikumi Katano, and Mamoru Ito. "Current advances in humanized mouse models." *Cellular & molecular immunology* 9, no. 3 (2012): 208.

Jacks, Tyler, Lee Remington, Bart O. Williams, Earlene M. Schmitt, Schlomit Halachmi, Roderick T. Bronson, and Robert A. Weinberg. "Tumor spectrum analysis in p53-mutant mice." *Current biology* 4, no. 1 (1994): 1-7.

Jiang, Le, Ning Kon, Tongyuan Li, Shang-Jui Wang, Tao Su, Hanina Hibshoosh, Richard Baer, and Wei Gu. "Ferroptosis as a p53-mediated activity during tumour suppression." *Nature*520, no. 7545 (2015): 57.

Joerger, A. C., and A. R. Fersht. "Structure–function–rescue: the diverse nature of common p53 cancer mutants." *Oncogene* 26, no. 15 (2007): 2226.

Joerger, Andreas C., and Alan R. Fersht. "Structural biology of the tumor suppressor p53." *Annu. Rev. Biochem.* 77 (2008): 557-582.

Joerger, Andreas C., and Alan R. Fersht. "The p53 pathway: origins, inactivation in cancer, and emerging therapeutic approaches." *Annual review of biochemistry* 85 (2016): 375-404.

Joerger, Andreas C., and Alan R. Fersht. "The tumor suppressor p53: from structures to drug discovery." *Cold Spring Harbor perspectives in biology* 2, no. 6 (2010): a000919.

Kaiser, Alyssa M., and Laura D. Attardi. "Deconstructing networks of p53-mediated tumor suppression in vivo." *Cell death and differentiation* 25, no. 1 (2018): 93.

Kalimutho, Murugan, Katia Nones, Sriganesh Srihari, Pascal HG Duijf, Nicola Waddell, and Kum Kum Khanna. "Patterns of genomic instability in breast cancer." *Trends in pharmacological sciences* (2019).

Kalluri, Raghu, and Robert A. Weinberg. "The basics of epithelial-mesenchymal transition." *The Journal of clinical investigation* 119, no. 6 (2009): 1420-1428.

Kandioler-Eckersberger, Daniela, Carmen Ludwig, Margarethe Rudas, Sonja Kappel, Elisabeth Janschek, Catharina Wenzel, Hermine Schlagbauer-Wadl et al. "TP53 mutation and p53 overexpression for prediction of response to neoadjuvant treatment in breast cancer patients." *Clinical Cancer Research*6, no. 1 (2000): 50-56.

Kang, Hua, Astrid Escudero-Esparza, Anthony Douglas-Jones, Robert E. Mansel, and Wen G. Jiang. "Transcript analyses of stromal cell derived factors (SDFs): SDF-2, SDF-4 and SDF-5 reveal a different pattern of expression and prognostic association in human breast cancer." *International journal of oncology* 35, no. 1 (2009): 205-211.

Kenny, Paraic A., and Mina J. Bissell. "Tumor reversion: correction of malignant behavior by microenvironmental cues." *International journal of cancer* 107, no. 5 (2003): 688-695.

Kim, Charissa, Ruli Gao, Emi Sei, Rachel Brandt, Johan Hartman, Thomas Hatschek, Nicola Crosetto, Theodoros Foukakis, and Nicholas E. Navin. "Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing." *Cell* 173, no. 4 (2018): 879-893.

Kim, Gwangil, Maria Ouzounova, Ahmed A. Quraishi, April Davis, Nader Tawakkol, Shawn G. Clouthier, Fayaz Malik et al. "SOCS3-mediated regulation of inflammatory cytokines in PTEN and p53 inactivated triple negative breast cancer model." *Oncogene* 34, no. 6 (2015): 671.

Kim, Jongdoo, Seunghee Bae, Sungkwan An, Jong Kuk Park, Eun Mi Kim, Sang-Gu Hwang, Wun-Jae Kim, and Hong-Duck Um. "Cooperative actions of p21WAF1 and p53 induce Slug protein degradation and suppress cell invasion." *EMBO reports* 15, no. 10 (2014): 1062-1068.

Kim, Michael P., and Guillermina Lozano. "Mutant p53 partners in crime." *Cell death and differentiation* 25, no. 1 (2018): 161.

Klebanov, Nikolai, Mykyta Artomov, William B. Goggins, Emma Daly, Mark J. Daly, and Hensin Tsao. "Burden of unique and low prevalence somatic mutations correlates with cancer survival." *Scientific reports* 9, no. 1 (2019): 4848.

Korkmaz, Gozde, Rui Lopes, Alejandro P. Ugalde, Ekaterina Nevedomskaya, Ruiqi Han, Ksenia Myacheva, Wilbert Zwart, Ran Elkon, and Reuven Agami. "Functional genetic screens for

enhancer elements in the human genome using CRISPR-Cas9." *Nature biotechnology* 34, no. 2 (2016): 192.

Korkmaz, Gozde, Rui Lopes, Alejandro P. Ugalde, Ekaterina Nevedomskaya, Ruiqi Han, Ksenia Myacheva, Wilbert Zwart, Ran Elkon, and Reuven Agami. "Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9." *Nature biotechnology* 34, no. 2 (2016): 192.

Kuperwasser, Charlotte, Gregory D. Hurlbut, Frances S. Kittrell, Ellen S. Dickinson, Rudy Laucirica, Daniel Medina, Stephen P. Naber, and D. Joseph Jerry. "Development of spontaneous mammary tumors in BALB/c p53 heterozygous mice: a model for Li-Fraumeni syndrome." *The American journal of pathology* 157, no. 6 (2000): 2151-2159.

Land, Hartmut, Luis F. Parada, and Robert A. Weinberg. "Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes." *Nature* 304, no. 5927 (1983): 596.

Lane, David P., and Lionel V. Crawford. "T antigen is bound to a host protein in SY40-transformed cells." *Nature* 278, no. 5701 (1979): 261.

Lang, Gene A., Tomoo Iwakuma, Young-Ah Suh, Geng Liu, V. Ashutosh Rao, John M. Parant, Yasmine A. Valentin-Vega et al. "Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome." *Cell* 119, no. 6 (2004): 861-872.

Langmead, Ben, and Steven L. Salzberg. "Fast gapped-read alignment with Bowtie 2." *Nature methods* 9, no. 4 (2012): 357.

Laptenko, O., and C. Prives. "Transcriptional regulation by p53: one protein, many possibilities." *Cell death and differentiation* 13, no. 6 (2006): 951.

Laptenko, Oleg, Idit Shiff, Will Freed-Pastor, Andrew Zupnick, Melissa Mattia, Ella Freulich, Inbal Shamir et al. "The p53 C terminus controls site-specific DNA binding and promotes structural changes within the central DNA binding domain." *Molecular cell* 57, no. 6 (2015): 1034-1046.

Lee, Jie-Oh, Haijuan Yang, Maria-Magdalena Georgescu, Antonio Di Cristofano, Tomohiko Maehama, Yigong Shi, Jack E. Dixon, Pier Pandolfi, and Nikola P. Pavletich. "Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association." *Cell* 99, no. 3 (1999): 323-334.

Lee, Yu-Ru, Ming Chen, and Pier Paolo Pandolfi. "The functions and regulation of the PTEN tumour suppressor: new modes and prospects." *Nature Reviews Molecular Cell Biology*(2018): 1.
Levine, Arnold J., and Moshe Oren. "The first 30 years of p53: growing ever more complex." *Nature reviews cancer* 9, no. 10 (2009): 749.

Levine, Arnold J., Jamil Momand, and Cathy A. Finlay. "The p53 tumour suppressor gene." *nature* 351, no. 6326 (1991): 453.

Li, Andrew G., Landon G. Piluso, Xin Cai, Gang Wei, William R. Sellers, and Xuan Liu. "Mechanistic insights into maintenance of high p53 acetylation by PTEN." *Molecular cell*23, no. 4 (2006): 575-587.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. "The sequence alignment/map format and SAMtools." *Bioinformatics* 25, no. 16 (2009): 2078-2079.

Li, Jie, Lixin Yang, Shikha Gaur, Keqiang Zhang, Xiwei Wu, Yate-Ching Yuan, Hongzhi Li, Shuya Hu, Yaguang Weng, and Yun Yen. "Mutants TP 53 p. R273H and p. R273C but not p. R273G Enhance Cancer Cell Malignancy." *Human mutation*35, no. 5 (2014): 575-584.

Liao, D. J., and R. B. Dickson. "c-Myc in breast cancer." *Endocrine-related cancer* 7, no. 3 (2000): 143-164.

Linzer, Daniel IH, and Arnold J. Levine. "Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells." *Cell* 17, no. 1 (1979): 43-52.

Linzer, Daniel IH, Warren Maltzman, and Arnold J. Levine. "The SV40 A gene product is required for the production of a 54,000 MW cellular tumor antigen." *Virology* 98, no. 2 (1979): 308-318.

Liu, D. P., Hoseok Song, and Yang Xu. "A common gain of function of p53 cancer mutants in inducing genetic instability." *Oncogene* 29, no. 7 (2010): 949.

Liu, Geng, Timothy J. McDonnell, Roberto Montes de Oca Luna, Mini Kapoor, Betsy Mims, Adel K. El-Naggar, and Guillermina Lozano. "High metastatic potential in mice inheriting a targeted p53 missense mutation." *Proceedings of the National Academy of Sciences* 97, no. 8 (2000): 4174-4179.

Liu, Jeff C., Veronique Voisin, Sharon Wang, Dong-Yu Wang, Robert A. Jones, Alessandro Datti, David Uehling et al. "Combined deletion of Pten and p53 in mammary epithelium accelerates triple-negative breast cancer with dependency on eEF2K." *EMBO molecular medicine* 6, no. 12 (2014): 1542-1560.

Liu, Xiaoling, Henne Holstege, Hanneke van der Gulden, Marcelle Treur-Mulder, John Zevenhoven, Arno Velds, Ron M. Kerkhoven et al. "Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1-mutated basal-like breast cancer." *Proceedings of the National Academy of Sciences* 104, no. 29 (2007): 12111-12116.

Lowe, Scott W., Stephan Bodis, Andrea McClatchey, Lee Remington, H. Earl Ruley, David E. Fisher, David E. Housman, and Tyler Jacks. "p53 status and the efficacy of cancer therapy in vivo." *Science* 266, no. 5186 (1994): 807-810.

Marchese, Stephanie, and Elisabete Silva. "Disruption of 3D MCF-12A breast cell cultures by estrogens–an in vitro model for ER-mediated changes indicative of hormonal carcinogenesis." *PloS one* 7, no. 10 (2012): e45767.

Martincorena, Iñigo, and Peter J. Campbell. "Somatic mutation in cancer and normal cells." *Science* 349, no. 6255 (2015): 1483-1489.

Martincorena, Iñigo, Keiran M. Raine, Moritz Gerstung, Kevin J. Dawson, Kerstin Haase, Peter Van Loo, Helen Davies, Michael R. Stratton, and Peter J. Campbell. "Universal patterns of selection in cancer and somatic tissues." *Cell* 171, no. 5 (2017): 1029-1041.

Marusyk, Andriy, Vanessa Almendro, and Kornelia Polyak. "Intra-tumour heterogeneity: a looking glass for cancer?." *Nature Reviews Cancer* 12, no. 5 (2012): 323.

Masciari, Serena, Deborah A. Dillon, Michelle Rath, Mark Robson, Jeffrey N. Weitzel, Judith Balmana, Stephen B. Gruber et al. "Breast cancer phenotype in women with TP53 germline mutations: a Li-Fraumeni syndrome consortium effort." *Breast cancer research and treatment* 133, no. 3 (2012): 1125-1130.

Mayer, Ingrid A., Vandana G. Abramson, Brian D. Lehmann, and Jennifer A. Pietenpol. "New strategies for triple-negative breast cancer—deciphering the heterogeneity." *Clinical cancer research* 20, no. 4 (2014): 782-790.

Mayo, Lindsey D., and David B. Donner. "The PTEN, Mdm2, p53 tumor suppressor–oncoprotein network." *Trends in biochemical sciences* 27, no. 9 (2002): 462-467.
McGranahan, Nicholas, and Charles Swanton. "Biological and therapeutic impact of intratumor heterogeneity in cancer evolution." *Cancer cell* 27, no. 1 (2015): 15-26.

Mello, Stephano S., and Laura D. Attardi. "Deciphering p53 signaling in tumor suppression." *Current opinion in cell biology*51 (2018): 65-72.

Mello, Stephano Spano, and Laura D. Attardi. "Not all p53 gain-of-function mutants are created equal." (2013): 855.

Miyamoto, Tsutomu, Akihisa Suzuki, Ryoichi Asaka, Kaori Ishikawa, Yasushi Yamada, Hisanori Kobara, Jun Nakayama, and Tanri Shiozawa. "Immunohistochemical expression of core 2 β1, 6-N-acetylglucosaminyl transferase 1 (C 2 G n T 1) in endometrioid-type endometrial carcinoma: a novel potential prognostic factor." *Histopathology* 62, no. 7 (2013): 986-993.
Muller, Patricia AJ, and Karen H. Vousden. "p53 mutations in cancer." *Nature cell biology* 15, no. 1 (2013): 2.

Muller, Patricia AJ, Karen H. Vousden, and Jim C. Norman. "p53 and its mutants in tumor cell migration and invasion." *The Journal of cell biology* 192, no. 2 (2011): 209-218.

Nik-Zainal, Serena, Helen Davies, Johan Staaf, Manasa Ramakrishna, Dominik Glodzik, Xueqing Zou, Inigo Martincorena et al. "Landscape of somatic mutations in 560 breast cancer whole-genome sequences." *Nature* 534, no. 7605 (2016): 47.

Nishida, Kozo, Keiichiro Ono, Shigehiko Kanaya, and Koichi Takahashi. "KEGGscape: a Cytoscape app for pathway data integration." *F1000Research* 3 (2014).

Norberg, Torbjörn, Sigrid Klaar, Gunilla Kärf, Hans Nordgren, Lars Holmberg, and Jonas Bergh. "Increased p53 mutation frequency during tumor progression—results from a breast cancer cohort." *Cancer research* 61, no. 22 (2001): 8317-8321.

Nowell, Peter C. "The clonal evolution of tumor cell populations." *Science* 194, no. 4260 (1976): 23-28.

Olive, Kenneth P., David A. Tuveson, Zachary C. Ruhe, Bob Yin, Nicholas A. Willis, Roderick T. Bronson, Denise Crowley, and Tyler Jacks. "Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome." *Cell* 119, no. 6 (2004): 847-860.

Olivier, Magali, Anita Langer, Patrizia Carrieri, Jonas Bergh, Sigrid Klaar, Jorunn Eyfjord, Charles Theillet et al. "The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer." *Clinical cancer research* 12, no. 4 (2006): 1157-1167.

Olivier, Magali, David E. Goldgar, Nayanta Sodha, Hiroko Ohgaki, Paul Kleihues, Pierre Hainaut, and Rosalind A. Eeles. "Li-Fraumeni and related syndromes: correlation between tumor type, family structure, and TP53 genotype." *Cancer research* 63, no. 20 (2003): 6643-6650.

Oren, Moshe, and Varda Rotter. "Mutant p53 gain-of-function in cancer." *Cold Spring Harbor perspectives in biology* 2, no. 2 (2010): a001107.

Parada, Luis F., Hartmut Land, Robert A. Weinberg, David Wolf, and Varda Rotter. "Cooperation between gene encoding p53 tumour antigen and ras in cellular transformation." *Nature*312, no. 5995 (1984): 649.

Pereira, Bernard, Suet-Feung Chin, Oscar M. Rueda, Hans-Kristian Moen Vollan, Elena Provenzano, Helen A. Bardwell, Michelle Pugh et al. "The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes." *Nature communications* 7 (2016): 11479.

Petitjean, Audrey, Ewy Mathe, Shunsuke Kato, Chikashi Ishioka, Sean V. Tavtigian, Pierre Hainaut, and Magali Olivier. "Impact of mutant p53 functional properties on TP53 mutation

patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database." *Human mutation*28, no. 6 (2007): 622-629.

Platt, Randall J., Sidi Chen, Yang Zhou, Michael J. Yim, Lukasz Swiech, Hannah R. Kempton, James E. Dahlman et al. "CRISPR-Cas9 knockin mice for genome editing and cancer modeling." *Cell* 159, no. 2 (2014): 440-455.

Polyak, Kornelia. "Heterogeneity in breast cancer." *The Journal of clinical investigation* 121, no. 10 (2011): 3786-3788.

Pon, Julia R., and Marco A. Marra. "Driver and passenger mutations in cancer." *Annual Review of Pathology: Mechanisms of Disease* 10 (2015): 25-50.

Powell, Emily, David Piwnica-Worms, and Helen Piwnica-Worms. "Contribution of p53 to metastasis." *Cancer discovery*4, no. 4 (2014): 405-414.

Ran, F. Ann, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. "Genome engineering using the CRISPR-Cas9 system." *Nature protocols* 8, no. 11 (2013): 2281.

Rangarajan, Annapoorni, Sue J. Hong, Annie Gifford, and Robert A. Weinberg. "Species-and cell type-specific requirements for cellular transformation." *Cancer cell* 6, no. 2 (2004): 171-183.

Rao, Donald D., John S. Vorhies, Neil Senzer, and John Nemunaitis. "siRNA vs. shRNA: similarities and differences." *Advanced drug delivery reviews* 61, no. 9 (2009): 746-759.

Rheinbay, Esther, Prasanna Parasuraman, Jonna Grimsby, Grace Tiao, Jesse M. Engreitz, Jaegil Kim, Michael S. Lawrence et al. "Recurrent and functional regulatory mutations in breast cancer." *Nature* 547, no. 7661 (2017): 55.

Rivlin, Noa, Ran Brosh, Moshe Oren, and Varda Rotter. "Mutations in the p53 tumor suppressor gene: important milestones at the various steps of tumorigenesis." *Genes & cancer* 2, no. 4 (2011): 466-474.

Roger, Lauréline, Laurent Jullien, Véronique Gire, and Pierre Roux. "Gain of oncogenic function of p53 mutants regulates E-cadherin expression uncoupled from cell invasion in colon cancer cells." *J Cell Sci* 123, no. 8 (2010): 1295-1305.

Rotter, Varda, Owen N. Witte, Robert Coffman, and David Baltimore. "Abelson murine leukemia virus-induced tumors elicit antibodies against a host cell protein, P50." *Journal of virology* 36, no. 2 (1980): 547-555.

Rotter, Varda. "p53, a transformation-related cellular-encoded protein, can be used as a biochemical marker for the detection of primary mouse tumor cells." *Proceedings of the National Academy of Sciences* 80, no. 9 (1983): 2613-2617.

Sansal, Isabelle, and William R. Sellers. "The biology and clinical relevance of the PTEN tumor suppressor pathway." *Journal of clinical oncology* 22, no. 14 (2004): 2954-2963.

Santoro, Angela, Thalia Vlachou, Lucilla Luzi, Giorgio Melloni, Luca Mazzarella, Errico D'Elia, Xieraili Aobuli et al. "p53 Loss in Breast Cancer Leads to Myc Activation, Increased Cell Plasticity, and Expression of a Mitotic Signature with Prognostic Value." *Cell reports* 26, no. 3 (2019): 624-638.

Sarnow, Peter, Ye Shih Ho, Jim Williams, and Arnold J. Levine. "Adenovirus E1b-58kd tumor antigen and SV40 large tumor antigen are physically associated with the same 54 kd cellular protein in transformed cells." *Cell* 28, no. 2 (1982): 387-394.

Sayols, Sergi, Denise Scherzinger, and Holger Klein. "dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data." *BMC bioinformatics* 17, no. 1 (2016): 428.

Scholl, Claudia, and Stefan Fröhling. "Exploiting rare driver mutations for precision cancer medicine." *Current opinion in genetics & development* 54 (2019): 1-6.

Sekine, Yuichi, Osamu Ikeda, Satoshi Tsuji, Chikako Yamamoto, Ryuta Muromoto, Asuka Nanbo, Kenji Oritani, Akihiko Yoshimura, and Tadashi Matsuda. "Signal-transducing adaptor protein-2 regulates stromal cell-derived factor-1α-induced chemotaxis in T cells." *The Journal of Immunology*183, no. 12 (2009): 7966-7974.

Shah, Sohrab P., Andrew Roth, Rodrigo Goya, Arusha Oloumi, Gavin Ha, Yongjun Zhao, Gulisa Turashvili et al. "The clonal and mutational evolution spectrum of primary triple-negative breast cancers." *Nature* 486, no. 7403 (2012): 395.

Shalem, Ophir, Neville E. Sanjana, and Feng Zhang. "High-throughput functional genomics using CRISPR–Cas9." *Nature Reviews Genetics* 16, no. 5 (2015): 299.

Shalem, Ophir, Neville E. Sanjana, Ella Hartenian, Xi Shi, David A. Scott, Tarjei S. Mikkelsen, Dirk Heckl et al. "Genome-scale CRISPR-Cas9 knockout screening in human cells." *Science* 343, no. 6166 (2014): 84-87.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. "Cytoscape: a software environment for integrated models of biomolecular interaction networks." *Genome research* 13, no. 11 (2003): 2498-2504.

Silwal-Pandit, Laxmi, Hans Kristian Moen Vollan, Suet-Feung Chin, Oscar M. Rueda, Steven McKinney, Tomo Osako, David A. Quigley et al. "TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance." *Clinical Cancer Research* 20, no. 13 (2014): 3569-3580.

Solatycka, Alicja, Tomasz Owczarek, Friedrich Piller, Véronique Piller, Bartosz Pula, Lukasz Wojciech, Marzena Podhorska-Okolow, Piotr Dziegiel, and Maciej Ugorski. "MUC1 in human and murine mammary carcinoma cells decreases the expression of core 2 β1, 6-N-acetylglucosaminyltransferase and β-galactoside α2, 3-sialyltransferase." *Glycobiology* 22, no. 8 (2012): 1042-1054.

Song, Hoseok, Monica Hollstein, and Yang Xu. "p53 gain-of-function cancer mutants induce genetic instability by inactivating ATM." *Nature cell biology* 9, no. 5 (2007): 573.

Sørlie, Therese, Charles M. Perou, Robert Tibshirani, Turid Aas, Stephanie Geisler, Hilde Johnsen, Trevor Hastie et al. "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications." *Proceedings of the National Academy of Sciences* 98, no. 19 (2001): 10869-10874.

Srivastava, Shiv, Zhiqiang Zou, Kathleen Pirollo, William Blattner, and Esther H. Chang. "Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li–Fraumeni syndrome." *Nature* 348, no. 6303 (1990): 747.

St Hill, Catherine A., Mariya Farooqui, Gregory Mitcheltree, H. Evin Gulbahce, Jose Jessurun, Qing Cao, and Bruce Walcheck. "The high affinity selectin glycan ligand C2-O-sLe x and mRNA transcripts of the core 2 β-1, 6-N-acetylglusaminyltransferase (C2GnT1) gene are highly expressed in human colorectal adenocarcinomas." *BMC cancer* 9, no. 1 (2009): 79.

Stambolic, V., D. MacPherson, D. Sas, Y. Lin, B. Snow, Y. Jang, S. Benchimol, and T. W. Mak. "Regulation of PTEN transcription by p53." *Molecular cell* 8, no. 2 (2001): 317-325.

Stephens, Philip J., Patrick S. Tarpey, Helen Davies, Peter Van Loo, Chris Greenman, David C. Wedge, Serena Nik-Zainal et al. "The landscape of cancer genes and mutational processes in breast cancer." *Nature* 486, no. 7403 (2012): 400.

Stratton, Michael R., Peter J. Campbell, and P. Andrew Futreal. "The cancer genome." *Nature* 458, no. 7239 (2009): 719.

Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich et al. "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." *Proceedings of the National Academy of Sciences* 102, no. 43 (2005): 15545-15550.

Sullivan, Kelly D., Matthew D. Galbraith, Zdenek Andrysik, and Joaquin M. Espinosa. "Mechanisms of transcriptional regulation by p53." *Cell death and differentiation* 25, no. 1 (2018): 133.

Sun, Ruping, Zheng Hu, and Christina Curtis. "Big Bang tumor growth and clonal evolution." *Cold Spring Harbor perspectives in medicine* 8, no. 5 (2018): a028381.

Tait, Larry, Herbert D. Soule, and Jose Russo. "Ultrastructural and immunocytochemical characterization of an immortalized human breast epithelial cell line, MCF-10." *Cancer research*50, no. 18 (1990): 6087-6094.

Tan, B. S., K. H. Tiong, H. L. Choo, F. Fei-Lei Chung, L. W. Hii, S. H. Tan, I. KS Yap et al. "Mutant p53-R273H mediates cancer cell survival and anoikis resistance through AKT-dependent suppression of BCL2-modifying factor (BMF)." *Cell death & disease* 6, no. 7 (2015): e1826.

Tidow, Henning, Roberto Melero, Efstratios Mylonas, Stefan MV Freund, J. Guenter Grossmann, José María Carazo, Dmitri I. Svergun, Mikel Valle, and Alan R. Fersht. "Quaternary structures of tumor suppressor p53 and a specific p53–DNA complex." *Proceedings of the National Academy of Sciences*104, no. 30 (2007): 12324-12329.

Turajlic, Samra, Andrea Sottoriva, Trevor Graham, and Charles Swanton. "Resolving genetic heterogeneity in cancer." *Nature Reviews Genetics* (2019): 1.

Vaughan, Catherine A., Swati P. Deb, Sumitra Deb, and Brad Windle. "Preferred binding of gain-of-function mutant p53 to bidirectional promoters with coordinated binding of ETS1 and GABPA to multiple binding sites." *Oncotarget* 5, no. 2 (2014): 417.

Ventura, Andrea, David G. Kirsch, Margaret E. McLaughlin, David A. Tuveson, Jan Grimm, Laura Lintault, Jamie Newman, Elizabeth E. Reczek, Ralph Weissleder, and Tyler Jacks. "Restoration of p53 function leads to tumour regression in vivo." *Nature* 445, no. 7128 (2007): 661.

Vogelstein, Bert, and Kenneth W. Kinzler. "The multistep nature of cancer." *Trends in genetics* 9, no. 4 (1993): 138-141.

Wan, Lei, Sean S. Molloy, Laurel Thomas, Gseping Liu, Yang Xiang, Sheree Lynn Rybak, and Gary Thomas. "PACS-1 defines a novel gene family of cytosolic sorting proteins required for trans-Golgi network localization." *Cell* 94, no. 2 (1998): 205-216.

Wang, Jing, Gregory J. Hannon, and David H. Beach. "Cell biology: risky immortalization by telomerase." *Nature* 405, no. 6788 (2000): 755.

Wang, Shu-Ping, Wen-Lung Wang, Yih-Leong Chang, Chen-Tu Wu, Yu-Chih Chao, Shih-Han Kao, Ang Yuan et al. "p53 controls cancer cell invasion by inducing the MDM2-mediated degradation of Slug." *Nature cell biology* 11, no. 6 (2009): 694.

Wang, Yong, Jill Waters, Marco L. Leung, Anna Unruh, Whijae Roh, Xiuqing Shi, Ken Chen et al. "Clonal evolution in breast cancer revealed by single nucleus genome sequencing." *Nature* 512, no. 7513 (2014): 155.

Weaver, Alissa M. "Invadopodia: specialized cell structures for cancer invasion." *Clinical & experimental metastasis* 23, no. 2 (2006): 97-105.

Wells, Mark, Henning Tidow, Trevor J. Rutherford, Phineus Markwick, Malene Ringkjobing Jensen, Efstratios Mylonas, Dmitri I. Svergun, Martin Blackledge, and Alan R. Fersht. "Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain." *Proceedings of the National academy of Sciences* 105, no. 15 (2008): 5762-5767.

Whyte, Peter, Karen J. Buchkovich, Jonathan M. Horowitz, Stephen H. Friend, Margaret Raybuck, Robert A. Weinberg, and Ed Harlow. "Association between an oncogene and an anti-oncogene: the adenovirus E1A proteins bind to the retinoblastoma gene product." *Nature* 334, no. 6178 (1988): 124.

Wijnhoven, Susan WP, Ewoud N. Speksnijder, Xiaoling Liu, Edwin Zwart, Rudolf B. Beems, Esther M. Hoogervorst, Mirjam M. Schaap et al. "Dominant-negative but not gain-of-function effects of a p53. R270H mutation in mouse epithelium tissue after DNA damage." *Cancer research* 67, no. 10 (2007): 4648-4656.

Williams, Marc J., Andrea Sottoriva, and Trevor A. Graham. "Measuring Clonal Evolution in Cancer with Genomics." *Annual review of genomics and human genetics* 20 (2019).

Willis, Amy, Eun Joo Jung, Therese Wakefield, and Xinbin Chen. "Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes." *Oncogene* 23, no. 13 (2004): 2330.

Wolf, D. A. V. I. D., and V. A. R. D. A. Rotter. "Inactivation of p53 gene expression by an insertion of Moloney murine leukemia virus-like DNA sequences." *Molecular and cellular biology* 4, no. 7 (1984): 1402-1410.

Wolf, David, and Varda Rotter. "Major deletions in the gene encoding the p53 tumor antigen cause lack of p53 expression in HL-60 cells." *Proceedings of the National Academy of Sciences* 82, no. 3 (1985): 790-794.

Wolf, David, Nicholas Harris, and Varda Rotter. "Reconstitution of p53 expression in a nonproducer Ab-MuLV-transformed cell line by transfection of a functional p53 gene." *Cell* 38, no. 1 (1984): 119-126.

Wood, Laura D., D. Williams Parsons, Siân Jones, Jimmy Lin, Tobias Sjöblom, Rebecca J. Leary, Dong Shen et al. "The genomic landscapes of human breast and colorectal cancers." *Science* 318, no. 5853 (2007): 1108-1113.

Yamaguchi, Hideki, Jeffrey Wyckoff, and John Condeelis. "Cell migration in tumors." *Current opinion in cell biology* 17, no. 5 (2005): 559-564.

Yates, Lucy R., Moritz Gerstung, Stian Knappskog, Christine Desmedt, Gunes Gundem, Peter Van Loo, Turid Aas et al. "Subclonal diversification of primary breast cancer revealed by multiregion sequencing." *Nature medicine* 21, no. 7 (2015): 751.

Yates, Lucy R., Stian Knappskog, David Wedge, James HR Farmery, Santiago Gonzalez, Inigo Martincorena, Ludmil B. Alexandrov et al. "Genomic evolution of breast cancer metastasis and relapse." *Cancer Cell* 32, no. 2 (2017): 169-184.

Yusuf, Rita, and Krystyna Frenkel. "Morphologic transformation of human breast epithelial cells MCF-10A: dependence on an oxidative microenvironment and estrogen/epidermal growth factor receptors." *Cancer Cell International* 10, no. 1 (2010): 30.

Zardavas, Dimitrios, Alexandre Irrthum, Charles Swanton, and Martine Piccart. "Clinical management of breast cancer heterogeneity." *Nature reviews Clinical oncology* 12, no. 7 (2015): 381.

Zhang, Yanhong, Wensheng Yan, and Xinbin Chen. "Mutant p53 disrupts MCF-10A cell polarity in three-dimensional culture via epithelial-to-mesenchymal transitions." *Journal of Biological Chemistry* 286, no. 18 (2011): 16218-16228.

Zhang, Yong, Tao Liu, Clifford A. Meyer, Jérôme Eeckhoute, David S. Johnson, Bradley E. Bernstein, Chad Nusbaum et al. "Model-based analysis of ChIP-Seq (MACS)." *Genome biology* 9, no. 9 (2008): R137.

Zhang, Yun, Shunbin Xiong, Bin Liu, Vinod Pant, Francis Celii, Gilda Chau, Ana C. Elizondo-Fraire et al. "Somatic Trp53 mutations differentially drive breast cancer and evolution of metastases." *Nature communications* 9, no. 1 (2018): 3953.

Zhao, Yahui, Xiaoding Hu, Li Wei, Dan Song, Juanjuan Wang, Lifang You, Hexige Saiyin et al. "PARP10 suppresses tumor metastasis through regulation of Aurora A activity." *Oncogene*(2018): 1.

Zheng, Tongsen, Jiabei Wang, Yuhan Zhao, Cen Zhang, Meihua Lin, Xiaowen Wang, Haiyang Yu, Lianxin Liu, Zhaohui Feng, and Wenwei Hu. "Spliced MDM2 isoforms

promote mutant p53 accumulation and gain-of-function in tumorigenesis." *Nature communications* 4 (2013): 2996.

Zhu, Jiajun, Morgan A. Sammons, Greg Donahue, Zhixun Dou, Masoud Vedadi, Matthäus Getlik, Dalia Barsyte-Lovejoy et al. "Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth." *Nature* 525, no. 7568 (2015): 206.

Zhu, Shiyou, Wei Li, Jingze Liu, Chen-Hao Chen, Qi Liao, Ping Xu, Han Xu et al. "Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library." *Nature biotechnology* 34, no. 12 (2016): 1279.

Zilfou, Jack T., and Scott W. Lowe. "Tumor suppressive functions of p53." *Cold Spring Harbor perspectives in biology*1, no. 5 (2009): a001883.

APPENDIX A

GUIDE RNA TARGETING PTEN GENE

Name	Spacer sequence	Strand	Transcript	Amino acid cut position	% peptide	Annotation	Score	Analysis from Broad institute
gRNA 1	GGTGGGTTATGGTCTTCAAA	antisense	ENST00000371953	91	22.58	CDS	0.55	match in PTENP1
gRNA 2	ATATCACCACACACAGGTAA	antisense		248	61.54		0.78	match in PTENP1
	PTEN exon 5							
gRNA 3	AGAGGCCCTAGATTTCTATG	sense					0.71	match in PTENP1

				Off-target site				Analysis from CRISPR design
	sgRNA sequence	Guide quality	Guide score	UCSC gene	Locus	Score		
	PTEN exon 2							
gRNA 4	TTATCCAAACATTATTGCTA	mid	41	NR_023917	chr9:-33676538	100		no matches in PTENP1
	PTEN exon 5							
gRNA 5	ATGTGCATATTTATTACATC	high	54	NR_023917	chr9:-33676218	5.4		no matches in PTENP1
	PTEN exon 6							
gRNA 6	CTACCTGTTAAAGAATCATC	high	66	NR_023917	chr9:-33676086	4.2		no matches in PTENP1

APPENDIX B

SELECTED GENES FROM RNA SEQ OF TP53 MUTANTS FOR VALIDATION

Gene	Av. expression level in 3 (Y234C, H179R, G245S) most non-invasive cell lines (TPM)	Av expression level in 3 (R248W, Y220C, R273C) most invasive cell lines (TPM)	Correlation between invasion level and RNA expression level
STAP2	0	14	0.63
CYP4B1	0	12	0.44
PARP10	0	2.8	0.59
PLPPR2	13	1.5	-0.39
SREBF1	24	4	-0.87
PDK2	12	2	-0.48

APPENDIX C

SELECTED GUIDE RNAS FOR INDIVIDUAL VALIDATION

Cell line screened	Gene	Selected sgRNA for individual validation
Y234C	FGD6	GATGATTTCCTCCTCATCAG
	GCNT1	TAGTCGTCAGGTGTCCACCG
	ADAM15	TGCCCATCGGGTGACCTGCC
R273C	PACS1	GAGATCGTCCTTCCAGCTAG
	SDF4	CGAGTTCATCTCCCTGCCCG